Continuous Linear Programming: Theory, Algorithms and Applications

by

Xiao-Dong Luo

B.S. in Computer Science, Peking University at BeiJing, China (1989)

M.S. in Computer Engineering, McMaster University at Ontario, Canada (1991)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1995

Author ....
Department of Electrical Engineering and Computer Science
June 16th, 1995

Certified by ....
Dimitris Bertsimas
Professor of Operations Research
Thesis Supervisor

Accepted by ....
Thomas L. Magnanti
George Eastman Professor of Management Science
Codirector, Operations Research Center

Continuous Linear Programming: Theory, Algorithms and Applications

by

Xiao-Dong Luo

Submitted to the Department of Electrical Engineering and Computer Science
on June 16th, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

Motivated by problems arising in queueing networks, manufacturing systems and communication networks, we propose a new subclass of continuous linear programming problems, called state constrainted separated continuous linear programming problems (or SCSCLP in short). These problems describe time-dependent averages in such systems and can be used for the control of these systems in a nonstationary environment. We demonstrate in this thesis that these problems can be efficiently solved using mathematical programming techniques.

As in finite dimensional linear programming, we investigate the SCSCLP with the help of its dual problem. In addition to the usual continuous linear programming dual, we propose an alternative dual problem for SCSCLP. We develop a new algorithm called the Successive Quadratic Programming method for the SCSCLP.

The new algorithm discretizes the problem over time. But unlike other algorithms proposed in the literature, it varies the discretization and the control simultaneously. Based on the number of constant pieces allowed in the control, we develop a quadratic program with polyhedral constraints. Even though the quadratic program is generally not convex, we apply nonlinear programming techniques such as the Frank-Wolfe method and the Matrix Splitting algorithm to get a KKT point for the quadratic program. By gradually increasing (and occasionally decreasing) the number of pieces allowed in the control, we can get better and better approximations thus improving any feasible solution that is not globally optimal for the SCSCLP. By bounding the size of the quadratic programming problems we encounter, we prove the finite convergence of the new algorithm. We also derive the optimal solution structure and prove absence of a duality gap as byproducts of the new algorithm. These type of results (i.e., finite convergence, optimal solution structure and absence of a duality gap) were known only under much more restrictive assumptions.

We then apply the theory we developed to specific multiclass fluid queueing networks, such as scheduling a) a multiclass queueing system with feedback (Klimov's problem), b) a multiclass queue under separable quadratic costs and c) a single class tandem queueing network. For the first problem, we show the optimality of an index rule, which shows that the problem is solvable in polynomial time. For the second problem, we propose a dynamic index rule that solves the fluid control problem. For the third problem, we prove the existence of a polynomial size optimal solution for the problem, which shows the problem is in NP∩CO-NP, a strong indication of the existence of a polynomial time algorithm for the problem. For the fluid multiclass queueing networks with routing, we give simple necessary and sufficient condition for the network to be stabilizable.

We also apply the theory we developed to fluid telephone loss networks. For this special class of linear optimal control problem with state feedback and constraints, we show that the problem admits a piecewise constant optimal control solution when the service rates are independent of the origin and destination of the calls. This new structural result gives a heuristic algorithm for the problem. We give a closed form optimal solution for a two class single-link fluid loss network, which provides insights to both the optimal solution structure for general fluid telephone loss networks and the corresponding stochastic control problem.

We have implemented our algorithms using the *C* programming language. We test our new algorithms using standard test problems and problems from manufacturing systems, communication networks and telephone loss networks. Our computational results show that the new algorithms can solve large scale problems efficiently.

Thesis Supervisor: Dimitris Bertsimas
Title: Professor of Operations Research

Acknowledgments

I would like to express my great gratitude to Professor Dimitris Bertsimas, for his support, guidance and encouragement. As my thesis supervisor, he always had time to listen and discuss, with genuine interest and extraordinary patience, especially considering he has been supervising ten plus students and has many other obligations.

I am also grateful to Professor Dimitri Bertsekas and Professor Thomas Magnanti, for their enlightening suggestions and wonderful job they did in serving on my thesis committee.

I am deeply indebted to Professor Paul Tseng, who is both a friend and a senior colleague. He has been a continuous guidance and source of encouragement.

Over the years I have also received numerous help, advice and encouragement from many other people, such as my high school mathematics and physics teachers, my former advisors in China and in Canada — Professor W. Zhou and Dr. S. Qiao.

I am very lucky to finish my Ph.D. at the Operations Research Center of MIT. Its students, staff and faculty have made the Center such an enjoyable place to study and do research. They have made my stay at MIT a lifetime experience. I would like to thank Paulette, Laura, Michele, Cheryl for the work they have done for the Center. I would also like to thank Arni, Asbjoern, David, Elaine, Jim, Jose, Rafael, Ronald, Thalia, Yi and Zhihang for their friendship.

Sincere thanks go to my brother Dr. Zhi–Quan Luo. He paved my way from high school to college, from China to Canada and then from Canada to MIT. He has been my role model and has given me help in almost every aspect of my life. I would like to thank my little sister and my parents. It is their faith and unsurpassed love that has kept me sane and hard-working. I dedicate this thesis to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this information superhighway era, it is very important to understand and control congestion while maintaining throughput in many production, service, communication and transportation systems. The goal of a manager of such systems is to evaluate, optimize and ultimately design these systems with the help of computers. Real systems can be complicated and sometimes even intractable. Assumptions on the arrival and service processes are usually made to simplify their analysis, without sacrificing accuracy. Among various techniques developed to represent, evaluate and optimize real systems, we are particularly interested in this thesis in the dynamic flow theory (or fluid models). Fluid models describe a system using time-dependent averages. We will show in this thesis that fluid models often accurately reflect a system's behavior and can be used to control queueing networks, manufacturing systems and communication systems. We will also demonstrate that fluid models can be efficiently solved using mathematical programming techniques.

1.1 Motivating Applications

Many applications naturally lend themselves to the use of fluid models. We next examine some of these applications.

1.1.1 A Fluid Queueing Network

Consider a tandem queueing network. As shown in Figure 1-1, we have from left to right two machines in tandem. The boxes represent machines. Jobs arrive at machine 1 at rate of λ. After a job completes service at machine 1, it moves to machine 2. After a job completes service at machine 2, it exits the system. Machine i can operate at any rate in $[0, \mu_i]$, for $i = 1, 2$. The network has $x_i(0)$ jobs at machine i at time 0. Before it starts its service at machine i, a job waits in a queue associated with the machine. Jobs at machine i cost w_i per unit time. Our objective is to control the service rates of the two machines such that the total cost of the jobs waiting in the network is minimized over a time horizon $[0,\ T]$.

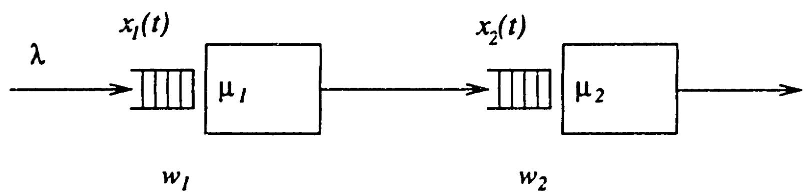

Figure 1-1: Two machines in tandem

We let $x_i(t)$ denote the number of jobs at machine i at time t. Obviously, we should have $x_i(t) \geq 0$ for all t. We let $u_i(t)$ denote the rate under which we operate machine i at time t. Obviously, they must satisfy the capacity constraints $0 \leq u_i(t) \leq \mu_i$ for all t. The problem can now be formulated as follows:

$$
\begin{array}{lll}
(TAND) & \min & \int_0^T [w_1 x_1(t) + w_2 x_2(t)]\, dt \\
& \text{such that} & \dot{x}_1(t) = \lambda - u_1(t) \\
& & \dot{x}_2(t) = u_1(t) - u_2(t) \\
& & 0 \leq u_1(t) \leq \mu_1 \\
& & 0 \leq u_2(t) \leq \mu_2
\end{array}
$$

$$x_1(t),\ x_2(t) \geq 0,$$

where $x(0)$ is a given vector.

By disregarding the randomness and focusing only on the dynamics of a stochastic queueing system, a fluid model often models the asymptotic behavior of the system under appropriate scaling. The use of the fluid model as an approximation for queueing systems is richly documented; see for example, Kleinrock [53], Newell [68], Hajek and Ogier [42], and numerous references therein. Chen and Mandelbaum [19], Chen [21], Dai [27] and Chen [22] showed that a wide range of queueing networks would converge (under appropriate time and space scaling) to fluid networks. In this sense, progress on optimizing fluid networks provides insights into queueing networks.

1.1.2 A Manufacturing System

We examine hierarchical production planning in a manufacturing system consisting of a network of machines (see Bai and Gershwin [8] and Sethi and Zhou [88]). The following example was introduced in [88].

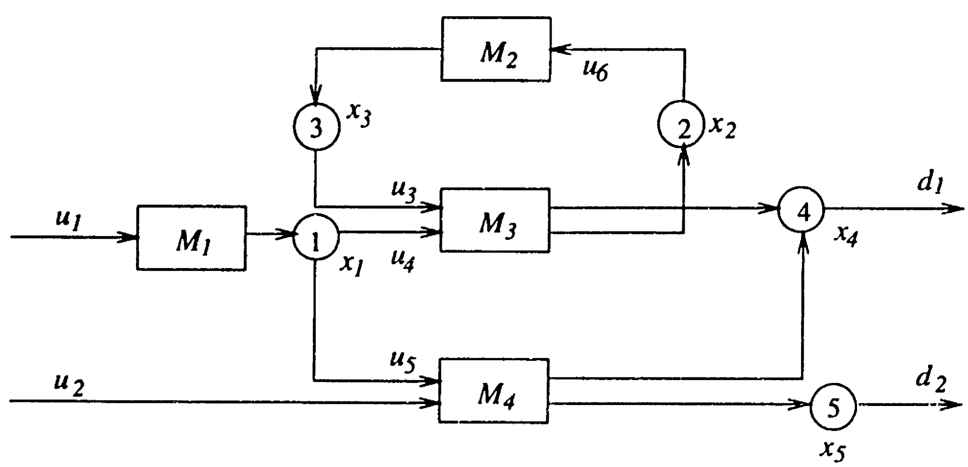

Figure 1-2: A typical manufacturing system

In Figure 1-2, we have four machines $M_1, \cdots, M_4$, two distinct products, and five buffers. The boxes in Figure 1-2 represent the machines and the circles represent the buffers. Each product $j = 1,2$ has demand d_j. As indicated in the figure, $x_i(t)$, $i = 1,2,\cdots,5$, is the state variable associated with buffer i. More specifically, $x_i(t)$ denotes the inventory/backlog of part type i at time t, $i = 1,2,\cdots,5$. Buffer 1 provides parts for machines 3 and 4, buffer 2 provides parts for machine 2, buffer 3 provides parts for machine 3, buffer 4 holds product 1 (i.e., part type 4) and buffer 5 holds product 2 (i.e., part type 5). Control variables $u_i(t), i = 1,2,\cdots,6$, represent the production rates. More specifically, $u_1(t)$ and $u_2(t)$ are the rates at which raw parts coming from the outside world are converted to part types 1 and 5, respectively, and $u_3(t), u_4(t), u_5(t)$ and $u_6(t)$ are the rates of conversion from part types $3,1,1$, and 2 to part types $4,2,4$ and 3, respectively. Therefore, the system dynamics associated with Figure 1-2 are

$$\begin{aligned} \dot{x}_1(t) &= u_1(t) - u_4(t) - u_5(t), \quad \dot{x}_3(t) = u_6(t) - u_3(t), \quad \dot{x}_5(t) = u_2(t) - d_2(t), \\ \dot{x}_2(t) &= u_4(t) - u_6(t), \quad \dot{x}_4(t) = u_3(t) + u_5(t) - d_1(t). \end{aligned} \tag{1.1}$$

As should be obvious from Figure 1-2 and the description above, part types 1, 2 and 3 are intermediate items to be further processed in the system. For $i = 1,2,3$, buffer i is between two machines and is considered an internal buffer. Internal buffers provide inputs to machines, and therefore must not have shortages, i.e., we must have

$$x(0) = x_0, \quad x_i(t) \geq 0, \qquad i = 1,2,3. \tag{1.2}$$

The remaining buffers 4 and 5 are called external buffers, since it is from these buffers that we must meet the demands of the final products facing the system. Since we permit backlogging of demand, the inventories in the external buffers can be negative. Indeed, $x_4(t)$ and $x_5(t)$ are called surpluses with positive values representing inventories and negative values representing backlogs. We denote $\mathcal{X}$ as the set of all feasible states $x(t)$ that satisfy both (1.1) and (1.2).

Each machine M_i, $i = 1,2,3,4$ has capacity $k_i(t)$ at time t. The control $u(t) = (u_1(t), \cdots, u_6(t))$ must satisfy the following capacity constraints

$$u_1(t) \leq k_1(t), \quad u_2(t) + u_5(t) \leq k_4(t),$$

$$u_3(t) + u_4(t) \leq k_3(t), \quad u_6(t) \leq k_2(t)$$
$$u_1(t), u_2(t), u_3(t), u_4(t), u_5(t), u_6(t) \geq 0. \tag{1.3}$$

A control $u(t)$ is admissible if it satisfies (1.3) for all $t \geq 0$. We use $\mathcal{U}$ to denote the set of all admissible controls.
We consider the discounted cost function $J(x(t), u(t))$ defined by

$$J(x(t), u(t)) = \int_0^T e^{-\rho t}[\sum_{i=1}^{5} h_i(x_i(t)) + c(u(t))]\, dt,$$

where $\rho > 0$ is the discount rate, T is the time horizon and $h_i(\cdot)$, representing the cost of surplus, is a piecewise linear function of the following form

$$h_i(x_i) = \begin{cases} \bar{h}_i \times x_i, & \text{if } x_i \geq 0; \\ -\tilde{h}_i \times x_i, & \text{if } x_i < 0. \end{cases}$$

Furthermore, $c(\cdot)$, which represents the cost of production, is a linear function. The problem is to find an admissible control $u(t)$ that minimizes $J(x(t), u(t))$. The problem can be written as follows:

$$\begin{array}{ll} \textit{minimize} & J(x(t), u(t)) = \int_0^T e^{-\rho t}[\sum_{i=1}^{5} h_i(x_i(t)) + c(u(t))]\, dt \\ \textit{subject to} & x(t) \in \mathcal{X}, u(t) \in \mathcal{U} \end{array}$$

which is equivalent to

$$\begin{array}{lll} (MANU) & \textit{minimize} & \int_0^T e^{-\rho t}[\bar{h}'x^+(t) + \tilde{h}'x^-(t) + c(u(t))]\, dt \\ & \textit{subject to} & (x^+(t) - x^-(t)) \in \mathcal{X}, u(t) \in \mathcal{U}, \\ & & x^+(t), x^-(t) \geq 0. \end{array}$$

In [88], Sethi and Zhou considered situations with four unreliable and failure prone machines. They assumed that the capacity $k_i(t)$ of machine i was governed by a finite state Markov process. They showed that as the rate at which the machines change states tends to infinity, the stochastic system is well approximated by $(MANU)$

with $k_i(t)$ being the average capacity for machine i. From any optimal solution to $(MANU)$, an asymptotically optimal control can be constructed for the stochastic system. The convergence rate as well as some computable error estimates of the value function of $(MANU)$ to the stochastic control problem can also be obtained.

1.1.3 Telephone Loss Networks

Fluid models are also useful for systems that process continuous flows, such as those in chemical and petroleum industries and in problems that have very small variability, such as telephone loss networks.

To demonstrate this, consider a telephone network defined on a complete digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V} = \{1, 2, \cdots, n\}$ and $\mathcal{E} = \{(i,j), i \neq j\}$ (see Figure 1-3). The telephone network consists of n different locations $i = 1, \cdots, n$ and $n \times (n-1)$ different links (i,j), for $i \neq j$. At time 0, there are some initial calls in the network. From each location i, calls to location j arise at a rate of λ_{ij}, while the duration of the calls is $\frac{1}{\mu_{ij}}$. Calls will either be accepted or rejected. If a call from i to j is accepted, it generates reward $\bar{w}_{ij}$ and can either be routed directly to location j through the link (i,j) that connects i and j, or be routed through (i,k) to a third location k and then from k to j through (k,j). If a call from i to j is rejected, a penalty of w_{ij} is incurred. We assume there are no other alternative routes and once a call is accepted, it cannot be interrupted. Every link (i,j) has a capacity of C_{ij} switching circuits. Every call consumes one switching circuit on every link it uses. Our goal is to decide whether to accept a call, and if we accept it, how we are going to route it, such that the sum of the weighted rewards of accepted calls less the penalty for lost calls is maximized over a period of time $[0, T]$.

To model this problem, we treat the number of calls as a function of time that takes on real values (i.e., as continuous flows) and formulate it in the following way. For any $i \neq j$ and $k \neq j$, we let $x_{ikj}(t)$ be the number of calls at time t in the network that are routed from location i to location j through location k. We use the convention that $x_{iij}(t)$ is the number of calls at time t that are routed directly from location i to location j. For any $i \neq j$ and $k \neq j$, we let $u_{ikj}(t)$ be the control variable that represents the rate at which calls made at time t from location i to location j are routed through location k. We use the convention that $u_{iij}(t)$ is the rate at which

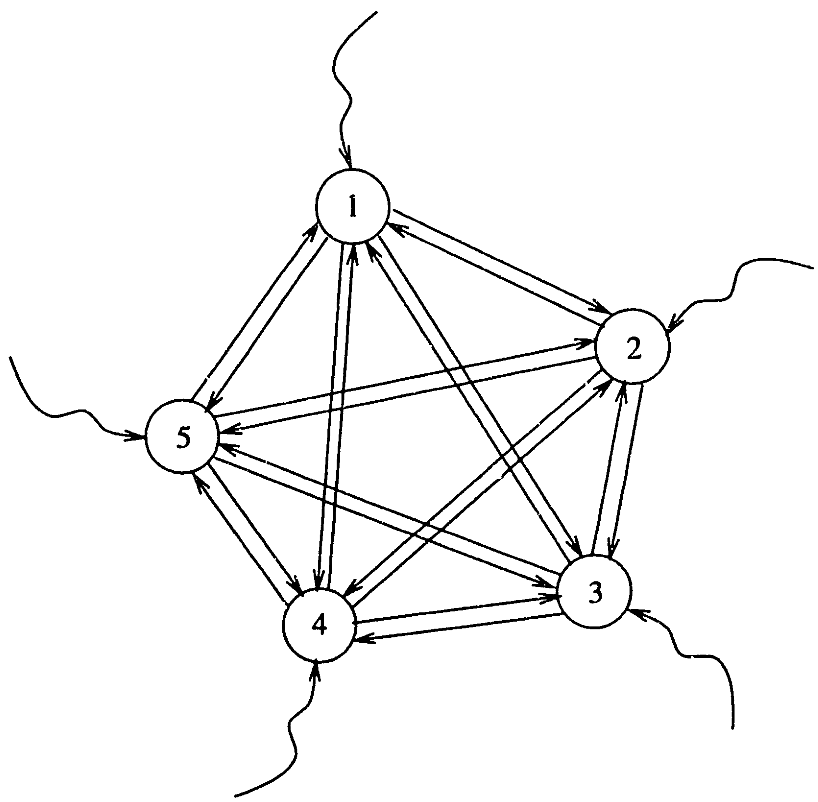

Figure 1-3: A five node telephone loss network

calls made at time t are routed directly from location i to location j. In this way, the variables $x_{ikj}(t)$ represent the state variables, while $u_{ikj}(t)$ represent the control variables. The problem can be formulated in a maximization form as follows:

$$\begin{aligned} (TLNa) \quad \textit{maximize} \quad & \int_0^T \left(\sum_{i,j} \bar{w}_{ij} \sum_k x_{ikj}(t) - \sum_{i,j} w_{ij} (\lambda_{ij} - \sum_k u_{ikj}(t)) \right) dt \\ \textit{subject to} \quad & x_{ikj}(t) = x_{ikj}(0) + \int_0^t (u_{ikj}(t) - \mu_{ij} x_{ikj}(t))\, dt, \qquad i \neq j,\ k \neq j \\ & \sum_{k \neq j} u_{ikj}(t) \leq \lambda_{ij}, \qquad i \neq j \end{aligned} \tag{1.4}$$

$$\sum_{k\neq j} x_{kij}(t) + \sum_{j\neq k, i\neq k} x_{ijk}(t) \leq C_{ij}, \qquad i \neq j \tag{1.5}$$
$$x(t),\ u(t) \geq 0, \qquad t \in [0,\ T],$$

where w and $\bar{w}$ are two nonnegative vectors.

If we assume $\mu_{ij} = \mu$ for all i and j, with a slight abuse of notation, we introduce new variables $y_{ikj}(t) = x_{ikj}(t)\ e^{\mu t}$ and define $u_{ikj}(t) = e^{\mu t}\ u_{ikj}(t)$. We now have the following equivalent problem:

$$\begin{aligned}
(TLNb) \quad & \textit{maximize} \quad \int_0^T \left(e^{-\mu t}\ w'u(t) + e^{-\mu t}\ \bar{w}'y(t)\right)\ dt \\
& \textit{subject to} \quad y_{ikj}(t) = y_{ikj}(0) + \int_0^t u_{ikj}(t)\ dt, \qquad i \neq j,\ k \neq j \\
& \qquad \sum_{k\neq j} u_{ikj}(t) \leq e^{\mu t}\lambda_{ij}, \qquad i \neq j \qquad (1.6) \\
& \qquad \sum_{k\neq j} y_{kij}(t) + \sum_{j\neq k, i\neq k} y_{ijk}(t) \leq e^{\mu t}\ C_{ij}, \qquad i \neq j \qquad (1.7) \\
& \qquad y(t),\ u(t) \geq 0, \qquad t \in [0,\ T].
\end{aligned}$$

Note that we can similarly formulated problems on undirected graphs.

1.2 Fluid Models and Continuous Linear Programs

A common characteristic for all the problems discussed in the previous section is that they all can be posed as linear fluid models defined as follows. A *linear fluid model* is a fluid model that can be formulated as the following optimization problem, first considered by Bellman [10]:

$$\begin{aligned}
(CLP) \quad & \textit{minimize} \quad \int_0^T c(t)'x(t)\ dt \\
& \textit{subject to} \quad A(t)x(t) + \int_0^t B(s,t)x(s)\ ds \leq b(t) \\
& \qquad x(t) \geq 0, \qquad t \in [0,\ T],
\end{aligned}$$

where $A(t)$ and $B(s,t)$ are matrices depending on time (their entries are bounded measurable functions) and $b(t)$ and $c(t)$ are bounded measurable functions. (CLP) is an instance of a *continuous linear program.*

The problem that attracted most attention (see Chapter 2) is the *separated continuous linear programming* problem, a subclass of the continuous linear programming problem:

$$\begin{aligned} (SCLP) \quad \text{minimize} \quad & \int_0^T c(t)'u(t)\, dt \\ \text{subject to} \quad & \int_0^t Gu(t)\, dt + y(t) = a(t) \qquad (1.8) \\ & Hu(t) \le b(t) \\ & y(t),\ u(t) \ge 0, \qquad t \in [0,\ T], \end{aligned}$$

where $y(t)$ and $a(t)$ are absolutely continuous functions. Note that the variables $u(t)$ and $y(t)$ are linked only through equation (1.8) where $u(t)$ appears only under the integration operator and $y(t)$ does not appear under the integration operator. The problem $(SCLP)$ was first introduced by Anderson in order to model job-shop scheduling problems [4]. It is easy to see that $(TAND)$ is an $(SCLP)$ if we change the constraints from differentiation to integration form.

In this thesis, we examine a larger subclass of continuous linear programs which contain the separated continuous linear programming problems and can be used to model a wider variety of problems that arise in communications, manufacturing and urban traffic control (e.g., $(MANU)$, $(TLNb)$). The problem we consider is the following:

$$\begin{aligned} (SCSCLP) \quad \text{minimize} \quad & \int_0^T (c(t)'u(t) + g(t)'y(t))\, dt \\ \text{subject to} \quad & \int_0^t Gu(t)\, dt + Ey(t) = a(t) \qquad (1.9) \\ & Hu(t) \le b(t) \qquad (1.10) \\ & Fy(t) \le h(t) \qquad (1.11) \\ & u(t) \ge 0, \qquad t \in [0,\ T], \end{aligned}$$

where $b(t)$, $c(t)$, $g(t)$ and $h(t)$ are bounded measurable functions and $a(t)$ is an absolutely continuous function. The dimensions of $b(t)$, $a(t)$, $u(t)$, $y(t)$ and $h(t)$ are n_1, n_2, n_3, n_4 and n_5, respectively. We call $(SCSCLP)$ the *state constrained separated continuous linear programs* (or SCSCLP in short). We call $y(t)$ the state variable and

$u(t)$ the control variable. We call (1.9) the state equation (or sometimes we use the term system dynamics) and call (1.11) the state constraint. We call (1.10) the control constraint.

Notice that there are three differences between $(SCLP)$ and $(SCSCLP)$. The first difference is in the objective function: $(SCLP)$ does not have a term containing the state variable while $(SCSCLP)$ does. The second difference is that the state equation of $(SCSCLP)$ contains a possibly nonsquare matrix E (the matrix is the identity matrix in $(SCLP)$). The third difference is that the state constraint of $(SCSCLP)$ contains a possibly nonsquare matrix F, while $(SCLP)$ only has non-negativity constraint imposed upon the state variable $y(t)$. $(SCLP)$ is obviously a special case of $(SCSCLP)$, obtained by setting E to the identity matrix, F to the negative indentity matrix, and $g(t)$ and $h(t)$ to zero vectors.

Some important applications cannot be modeled as $(SCLP)$ while they can be easily modeled as an $(SCSCLP)$. The problem $(MANU)$ is an $(SCSCLP)$, but not an $(SCLP)$. $(TLNb)$ is an $(SCSCLP)$, but not an $(SCLP)$. On the other hand, if the duration of a call $\frac{1}{\mu_{ij}}$ depends on the origin and destination, then $(TLNa)$ is not equivalent to $(TLNb)$ and $(TLNa)$ cannot be modeled as an $(SCSCLP)$.

When μ_{ij} depends on the origin and destination of the call, $(TLNa)$ is an instance of the linear optimal control problem, a fundamental problem in modern control theory. The *linear optimal control* problem is the following:

$$\begin{aligned}
(LCONT) \quad \textit{minimize} \quad & \int_0^T [c(t)'u(t) + g(t)'y(t)]\, dt \\
\textit{subject to} \quad & \dot{y}(t) = \dot{a}(t) - Gu(t) + Dy(t) \qquad (1.12) \\
& y(0) = a(0) \\
& u(t) \in \mathcal{U}(t) = \{u(t) : u(t) \geq 0,\ Hu(t) \leq b(t)\} \\
& y(t) \in \mathcal{Y}(t) = \{y(t) : Fy(t) \leq h(t)\}, \qquad t \in [0,\ T],
\end{aligned}$$

where $a(t)$, $b(t)$, $c(t)$, $g(t)$, $h(t)$ have the same restrictions as in $(SCSCLP)$. The term $Dy(t)$ in (1.12) is called the state feedback for the problem.

When $D = 0$, the problem $(LCONT)$ can be transformed into an equivalent $(SCSCLP)$ by formally integrating (1.12) over $[0,\ t]$ for all t. By differentiating the state constraint (1.6), $(SCLP)$ can be transformed into an equivalent $(LCONT)$.

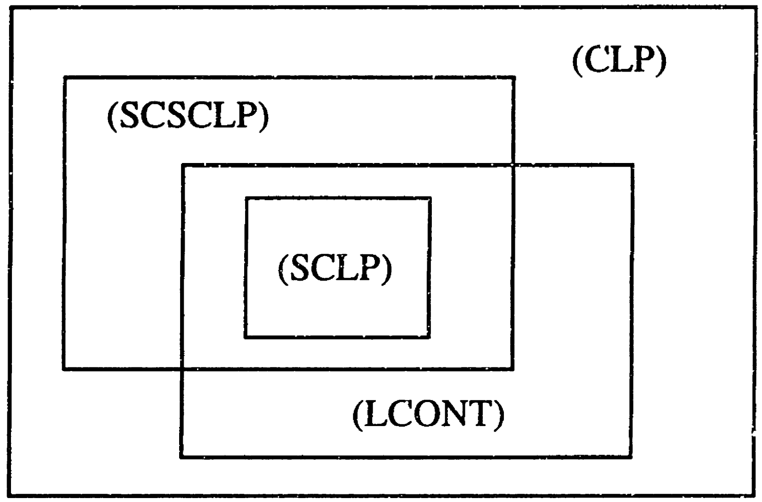

Figure 1-4: Hierarchy of different formulations

Integrating (1.12) over $[0, t]$ for all t shows that $(LCONT)$ is also a subclass of (CLP). The relation among $(SCLP)$, $(SCSCLP)$, $(LCONT)$ and (CLP) is illustrated in Figure 1-4.

1.3 Contribution and Structure of this Thesis

Motivated by the application areas reviewed in Section 1.1, our goal in this thesis is to develop and study a new algorithm for SCSCLP as well as for its special case SCLP. In particular, our contribution is to:

1) propose an efficient algorithm for solving large scale SCSCLP problems,

2) introduce a new dual problem for $(SCSCLP)$ and by using quadratic programming techniques, prove algorithmically the absence of a duality gap for SCSCLP without assuming that the primal solution set is bounded,

3) provide conditions for the existence of an optimal solution,

4) provide new optimal solution structural results for SCSCLP, and

5) apply the new algorithm to a variety of queueing control problems and problems arising from communication and manufacturing systems.

The rest of the thesis is structured as follows.

In Chapter 2, we review the general theory of continuous linear programs and the related literature, in order to make the thesis self-contained.

In Chapter 3, we propose a new numerical algorithm for SCSCLP, prove its convergence, and show the absence of a duality gap for the problem.

In Chapter 4, we use the theory developed in Chapters 2 and 3 to analyze several queueing control problems with particular structure. We also examine the controllability of multiclass queueing networks with routing.

In Chapter 5, we provide new structural results for problems arising from telephone loss networks and propose an algorithm for solving it. We also give a closed form solution to a single-link loss network problem that provides insights to both the optimal solution structure for general fluid telephone loss networks and the corresponding stochastic control problem.

In Chapter 6, we examine the computational behavior of the new algorithm and provide some numerical examples.

In Chapter 7, we give conclusions and pose some open questions.

Chapter 2

General Theory of CLP

In Chapter 1, we introduced fluid models and various continuous linear programs. In this chapter, we provide a historical review and some general results on continuous linear programming. We refer the reader to the book by Anderson and Nash [2] and the papers by Pullan [77, 78, 79] for more detailed results and analysis.

2.1 A Brief Historical Review

Bellman [10, 11] was the first person to introduce problem (CLP). Work on CLP is divided into three broad areas: the establishment of strong duality theorems, the development of algorithms, and the characterization of the optimal solution structure.

After problem (CLP) was introduced, a lot of papers dealing with duality theory appeared. Tyndall [92] gave conditions under which (CLP) and its dual have the same value, when they are both posed in L_∞ spaces. Levinson [58] extended Tyndall's work by considering time dependent matrices. So did Tyndall in [93]. In all this work, the strong duality theorem was proved using a sequence of successively finer discretizations. Grinold [38] used the abstract mathematical programming approach. Yamasaki [99] also used this approach and considered the problems posed over other functional spaces. Rockafellar [81] considered a class of state constrained convex optimal control problems. Hager and Mitter [40] and Hager [41] extended his work by considering a larger class of problems, which included ($LCONT$). They established a general duality theory under certain Slater conditions. Duality theory was also addressed by Papageorgiou [70] from a different perspective. In order to obtain a more satisfactory duality theory, he modified the objective functions and posed both the primal and the dual problems in the space of functions of bounded variation. Sethi *et. al.* [86] investigated the problem as an optimal control problem with inequality constraints on the state variables. They related the optimal dual variables to the adjoint variables in the optimal control problem.

The computational study of CLP was initiated by Lehman [57] who attempted to develop simplex-like algorithm for CLP. Drews [29], Hartberger [45] and Segers [84] later followed him. Perold [72] developed the first simplex-like algorithm for CLP. Anstreicher [7] continued Perold's work in his Ph.D. thesis, even though both of their algorithms were still incomplete and rather complicated. In the meantime, Russian authors such as Ilyutovich [46, 47] treated the problem using Pontryagin's Maximum Principle. In addition, Ito *et. al.* [48] has developed a primal-dual path following interior point method for CLP.

Perold [73] studied the extreme point structure for (CLP), under some mild assumptions that guarantee their existence. Along the same line, Anderson, Nash and Perold [1] derived extreme point structure for $(SCLP)$. Based on the optimal solution structure, Anderson and Nash in [2] proposed a convex quadratic programming procedure for $(SCLP)$.

Only recently, work on CLP that combined all of the above areas has appeared. The series of papers on SCLP by Pullan [77, 78, 79], deal with solution structure, duality theory and numerical algorithms. Philpott later specialized Pullan's work to a network version of SCLP [74] and presented encouraging numerical results.

2.2 Extreme Points and BFS

Throughout this thesis, unless otherwise specified, we always assume the primal problems (CLP), $(SCLP)$ and $(SCSCLP)$ are feasible. In this section, we review the notion of a basic feasible solution for these linear programs.

We call a feasible solution to problem (CLP) a basic feasible solution (or BFS in short) if it cannot be written as a convex combination of two distinct feasible solutions (different on a set of positive measure) to (CLP). As in finite dimensional linear programming, we also call a basic feasible solution an extreme point of the feasible region.

In finite dimensional linear programming, there always exists a basic feasible solution that is optimal for a linear program, given its feasible region is bounded. A similar result holds for $(SCLP)$. The following proposition comes from Anderson [5].

Proposition 2.1 *For the problem $(SCLP)$, if*

a) $a(t)$ is continuous and piecewise differentiable with a bounded derivative,

b) $u(t)$ is bounded by the constraints,

then there exists a basic feasible solution that is optimal to the problem $(SCLP)$.

For this reason, it is important to characterize the structure of the extreme points. In fact, the extreme points of $(SCLP)$ have a simple characterization. Some more notation is needed before giving this characterization. We define the matrix K as

$$K = \begin{pmatrix} G & I & 0 \\ H & 0 & I \end{pmatrix}.$$

For any feasible solution $x(t) = (u(t),\ y(t))$ to problem $(SCLP)$, let $z(t) = b(t) - Hu(t)$ and $\bar{x} = (u(t),\ y(t),\ z(t))$ and define $S_x(t)$ as the set valued function such that

$$S_x(t) = \{\ k \mid \bar{x}_k(t) \neq 0\ \}.$$

We call $S_x(t)$ the support function of $x(t)$. The following proposition also comes from Anderson [5].

Proposition 2.2 *A feasible solution $x(t) = (u(t),\ y(t))$ to problem $(SCLP)$ is a basic feasible solution if and only if the columns of K indexed by $S_x(t)$ are linearly independent almost everywhere in $[0,\ T]$.*

For general (CLP), Perold [73] characterized the extreme points that are right analytic. We call a function $f : [0,\ T] \mapsto \Re$ right analytic if for each $t \in [0,\ T)$, there is an ϵ and an analytic function $g : (t-\epsilon,\ t+\epsilon) \mapsto \Re$ such that $f(s) = g(s)$ for all $s \in [t,\ t+\epsilon)$. Not all the extreme points for (CLP) are right analytic. It is not known whether the optimal objective value is always attained at such an extreme point even if we assume there is an optimal basic feasible solution and the problem data are well behaved. Since Perold's characterization is not directly related to the thesis, we did not include it in the thesis.

A better characterization can be obtained if we insist that the basic feasible solution is also optimal. The following proposition is due to Anderson and Nash [2].

Proposition 2.3 *If the two conditions in Proposition 2.1 hold and $a(t)$, $c(t)$ are piecewise linear and $b(t)$ is piecewise constant, then there exists an optimal basic feasible solution to $(SCLP)$ whose $u(t)$ is piecewise constant.*

The above result was recently extended by Pullan [78] for the cases where $a(t)$, $b(t)$ and $c(t)$ are piecewise analytic. He showed that there exists an optimal feasible solution for $(SCLP)$ whose $u(t)$ is piecewise analytic assuming that the two conditions in Proposition 2.1 hold.

2.3 Duality Theory

As in finite dimensional linear programming, both theoretical and algorithmic developments rely heavily on the study of the dual problem. In this section, we introduce dual problems for (CLP), $(SCLP)$ and $(SCSCLP)$. We also provide several useful duality results for (CLP).

2.3.1 The Dual Problems

We take an abstract mathematical programming approach. Again, the theory developed here is not meant to be the most general possible. We refer the reader to Luenberger [60], Rockafellar [81] and Anderson and Nash [2] for more material on this subject.

Intentionally, when we introduced (CLP) and $(SCSCLP)$ in Chapter 1, we did not specify the functional spaces to which $x(t)$, $y(t)$ and $u(t)$ should belong. The specification of these spaces is very important, however, since the duality results that hold in one space might not hold in another. One contribution of the thesis is that we identify the spaces we need in order to establish strong duality results. First, we introduce the dual problem for the following problem, posed abstractly in functional spaces. Let (X, Y) and (Z, W) be two dual pairs of vector spaces.

$$\begin{aligned} (MP) \quad \textit{minimize} \quad & \langle x, c \rangle \\ \textit{subject to} \quad & (\mathcal{A}x - b) \in \mathcal{P}_2 \\ & x \in \mathcal{P}_1, \end{aligned}$$

where $\mathcal{A}$ is a linear operator that maps an element of X to an element in Z, b and c are given elements of Z and Y respectively. $\mathcal{P}_1$ and $\mathcal{P}_2$ are convex positive cones in X and Z respectively.

We introduce the dual problem of (MP):

$$\begin{aligned} (MP^*) \quad \textit{maximize} \quad & \langle b, w \rangle \\ \textit{subject to} \quad & -(\mathcal{A}^* w - c) \in \mathcal{P}_1^* \\ & w \in \mathcal{P}_2^* \end{aligned}$$

where $\mathcal{P}_i^*$ is the dual cone for $\mathcal{P}_i$ and $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$ defined by

$$\langle x, \mathcal{A}^* y \rangle = \langle \mathcal{A}x, y \rangle.$$

Let X be $L_\infty^{n_1}[0,\ T]$, Y be $L_1^{n_1}[0,\ T]$, Z be $L_1^{n_2}[0,\ T]$ and W be $L_\infty^{n_2}[0,\ T]$. We define the bilinear forms and the linear operator $\mathcal{A}$ as

$$\begin{aligned} \langle c(t), x(t) \rangle &= \int_0^T c(t)' x(t)\ dt \\ \langle b(t), w(t) \rangle &= \int_0^T b(t)' w(t)\ dt \\ \mathcal{A}x(t) &= A(t)x(t) + \int_0^t B(s,t)x(s)ds \end{aligned}$$

where $c(t)$ and $x(t)$ are in Y and X respectively, $b(t)$ and $w(t)$ are in Z and W respectively. $\mathcal{P}_1$ and $\mathcal{P}_2$ consist of vector functions with dimensions n_1 and n_2 respectively that are nonnegative a.e. on $[0,\ T]$. We can use the Fubini theorem to find the adjoint operator of $\mathcal{A}$. It is also easy to characterize $\mathcal{P}_1^*$ and $\mathcal{P}_2^*$ from their definitions. With the above abstract setting in mind, we can now introduce the dual of (CLP), which Bellman [11] devised when he introduced the (CLP).

$$
\begin{aligned}
(CLP^*) \quad & maximize && -\int_0^T b(t)'w(t)\,dt \\
& subject\ to && c(t) + A(t)'w(t) + \int_t^T B(s,t)'w(s)\,ds \geq 0 \\
& && w(t) \geq 0, \qquad \text{for } t \in [0,\ T],
\end{aligned}
$$

where $w(t)$ is a bounded measurable function.

Restricted to $(SCLP)$ (see also [2]), we can pose the following dual problem for $(SCLP)$:

$$
\begin{aligned}
(SCLP1^*) \quad & maximize && -\int_0^T \pi(t)'a(t)\,dt - \int_0^T b(t)'\eta(t)\,dt \\
& subject\ to && c(t) + \int_t^T G'\pi(t)\,dt + H'\eta(t) \geq 0 \\
& && \pi(t) \geq 0, \eta(t) \geq 0, \qquad \text{for } t \in [0,\ T],
\end{aligned}
$$

with $\pi(t)$ and $\eta(t)$ in the space of bounded measurable functions.

When restricted to $(SCSCLP)$, (MP^*) gives the following dual problem to $(SCSCLP)$:

$$
\begin{aligned}
(SCSCLP1^*) \quad & maximize && -\int_0^T a(t)'\pi(t)\,dt - \int_0^T b(t)'\eta(t)\,dt - \int_0^T h(t)'\xi(t)\,dt \\
& subject\ to && c(t) + \int_t^T G'\pi(t)\,dt + H'\eta(t) \geq 0 \\
& && E'\pi(t) + F'\xi(t) = -g(t) \\
& && \eta(t) \geq 0, \xi(t) \geq 0, \qquad \text{for } t \in [0,\ T].
\end{aligned}
$$

Again, the dual variables are bounded measurables functions.

2.3.2 Duality Results and Alternative Dual Problems

Duality theory is at the heart of the simplex method and many other recently developed barrier type methods for finite dimensional linear programming. To extend these algorithms to CLP (or SCLP and SCSCLP), it would be necessary to establish a similar duality theory for CLP (or SCLP and SCSCLP). This idea appeared in the early development of partial algorithms in Lehman [57], Drews [29], Hartberger [45], Segers [84], Perold [72] and Anstreicher [7]. Because of the importance of the duality results, many papers dealt solely with duality for CLP.

In introducing the problem, Bellman readily established the weak duality result between (CLP) and (CLP^*), i.e., the objective value of (CLP) is always no less than that of (CLP^*). This result is a special case of general weak duality results between (MP) and (MP^*) (see Anderson and Nash [2]). The first strong duality results for (CLP) and (CLP^*) were given by Tyndall [92]. Among other things, the strong duality result in [92] required that $A(t)$, $B(s,t)$ and $b(t)$ be nonnegative. Consequently, this result is not very useful as many practical problems (e.g. network problems in Anderson and Philpott [3]) give rise to negative entries in G and $a(t)$ in $(SCLP)$, and hence also in $B(s,t)$ and $b(t)$ in the corresponding (CLP).

Tyndall's work was extended by Levinson [58], Tyndall [93] and Grinold [37, 38, 39] (see also Anderson and Nash [2]). The problem of establishing strong duality results was not fully solved because many simple CLPs, such as the network version of SCLP, were not covered by their results. More general results have not been obtained, because it is not difficult to construct counter-examples that exhibit duality gaps for general (MP) and (MP^*). For example, in semidefinite programming, a finite duality gap can be present, even if both the primal and the dual problem attain their optimal solution values. Counter-examples were also constructed for (CLP), using the dual problem (CLP^*) (see, for example, Grinold [37]). We note that Grinold's example does not show the existence of a duality gap between (CLP) and (CLP^*). This means that to establish more general strong duality results (i.e., both the existence of the optimal solutions and the absence of a duality gap) for (CLP) (or $(SCLP)$ and $(SCSCLP)$), it is necessary to consider a dual problem in some sense different from (CLP^*). This was noted by Lehman [57], who tried to develop algorithms for (CLP). Rockafellar [81] noted that strong duality theorems could be obtained by allowing dual variables to be of bounded variation for a class of control problems. The establishment of a strong duality result was also addressed by Papageorgiou [70] from a different perspective. He modified the objective functions and posed both the primal problem and the dual problem in the space of functions of bounded variation. However, the new problem he studied generally had a different objective value than (CLP). His proof technique was not constructive and the connection between the optimal solution of his problem and the original problem (CLP) has not yet been established.

The idea of finding an alternative dual was most recently rediscovered by Pullan [77], who expanded the feasible solutions to the dual problem by allowing Dirac δ functions. However, the existence of an optimal solution does not follow from the boundedness of the objective value for $(SCLP)$, in contrast to finite dimensional linear programming. Assumptions that guarantee the existence of an optimal solution for the primal problem have been a common characteristic in the literature on duality theory for (CLP).

Pullan [77] introduced the following dual problem for $(SCLP)$:

$$\begin{array}{lll}(SCLP^*) & \textit{maximize} & -\int_0^T a(t)'\, d\pi(t) - \int_0^T b(t)'\eta(t)\, dt \\ & \textit{subject to} & c(t) - G'\pi(t) + H'\eta(t) \geq 0 \\ & & \eta(t) \geq 0, \qquad \text{for } t \in [0,\ T], \\ & & \pi(t) \text{ monotonic increasing and right continuous} \\ & & \text{on } [0,\ T] \text{ with } \pi(T) = 0.\end{array}$$

The basic idea is to expand the feasible region of $\pi(t)$ in $(SCLP1^*)$ to include the unbounded Dirac δ functions. To see this, we let $(\pi_1(t),\ \eta_1(t))$ be a feasible solution for $(SCLP1^*)$. If we define

$$\pi(t) = -\int_t^T \pi_1(t)\, dt, \qquad \eta(t) = \eta_1(t),$$

we see that $(\pi(t),\ \eta(t))$ is a feasible solution for $(SCLP^*)$ with the same objective value. However, unless $\pi(t)$ is absolutely continuous, the above process cannot be reversed due to Proposition A.1 in the Appendix. As we will see later in Chapter 3, there always exists an optimal solution to $(SCLP^*)$ with $\pi(t)$ piecewise absolutely continuous (cf. Theorem 3.9). By the definition of the δ function in the Appendix, this means that the only "useful" extensions made on $(SCLP1^*)$ is the inclusion of the δ functions at the points where $\pi(t)$ has jumps in the dual solution for $(SCLP^*)$. It is this type of extension that makes the proof of the absence of a duality gap in the subsequent chapters work.

The following proposition comes from Pullan [77].

Proposition 2.4 *Weak duality holds between* $(SCLP)$ *and* $(SCLP1^*)$ *and the optimal value of* $(SCLP^*)$ *lies between that of* $(SCLP)$ *and* $(SCLP1^*)$.

It is not known whether a duality gap exists between $(SCLP)$ and $(SCLP1^*)$. By utilizing Proposition 2.3, Pullan [77, Corollary 4.5] algorithmically showed the following:

Proposition 2.5 *Under the same assumption as in Proposition 2.3, there exist piecewise linear optimal solutions for both* $(SCLP)$ *and* $(SCLP^*)$; *furthermore, strong duality holds between* $(SCLP)$ *and* $(SCLP^*)$.

The above result has been recently strengthened by Pullan [79] who assumed that $a(t)$, $b(t)$ and $c(t)$ are piecewise analytic instead of piecewise linear or piecewise constant. However, the boundedness assumption on the feasible solution set for $(SCLP)$ is still imposed (cf. Assumption b in Proposition 2.1).

Throughout the remainder of this chapter and in Chapter 3, we make the following assumptions for problem $(SCSCLP)$:

Assumption 2.1

a) $a(t)$ *and* $h(t)$ *are continuous,*

b) $a(t)$, $c(t)$ *and* $h(t)$ *are piecewise linear,*

c) $b(t)$ *and* $g(t)$ *are piecewise constant,*

d) problem $(SCSCLP)$ *is feasible and its objective value is bounded from below.*

We prove our new duality results by introducing δ-functions in the dual problem and by using quadratic programming techniques to express possibly unbounded primal feasible solutions. However, we only explicitly write the dual problem in a different form. Similar to $(SCLP^*)$, we propose the following alternative dual problem for $(SCSCLP)$:

$$\begin{array}{lll}(SCSCLP^*) & \textit{maximize} & -\int_0^T a(t)'\, d\pi(t) - \int_0^T b(t)'\eta(t)\, dt - \int_0^T h(t)'\, d\xi(t) \\ & \textit{subject to} & c(t) - G'\pi(t)\, dt + H'\eta(t) \geq 0\end{array}$$

$$
\begin{aligned}
&E'\pi(t) + F'\xi(t) = \int_t^T g(t)\ dt \\
&\pi(t) \text{ is a bounded measurable VF function} \\
&\xi(t) \text{ monotonic increasing and right continuous} \\
&\quad \text{on } [0,\ T] \text{ with } \xi(T) = 0, \quad \pi(T) = 0 \\
&\eta(t) \geq 0, \qquad \text{for } t \in [0,\ T].
\end{aligned}
$$

Similar to Proposition 2.4, we have the following weak duality results for $(SCSCLP)$. For completeness, we give its proof.

Proposition 2.6 *Weak duality holds between* $(SCSCLP)$ *and* $(SCSCLP1^*)$ *and the optimal value of* $(SCSCLP^*)$ *lies between that of* $(SCSCLP)$ *and* $(SCSCLP1^*)$.

Proof It is clear that every feasible solution to $(SCSCLP1^*)$ corresponds to a feasible solution to $(SCSCLP^*)$ which has the same solution value. Therefore, the objective value of $(SCSCLP^*)$ is always greater than or equal to that of $(SCSCLP1^*)$. So, we only need to show that weak duality holds between $(SCSCLP)$ and $(SCSCLP^*)$.

Consider any two solutions $(u(t),\ y(t))$ and $(\pi(t),\ \eta(t),\ \xi(t))$ which are feasible to $(SCSCLP)$ and $(SCSCLP^*)$ respectively. Let $z(t) = b(t) - Hu(t)$. We have

$$
\begin{aligned}
& \int_0^T (c(t)'u(t) + g(t)'y(t))\ dt - \left(-\int_0^T a(t)'\ d\pi(t) - \int_0^T b(t)'\eta(t)\ dt - \int_0^T h(t)'\ d\xi(t)\right) \\
= & \int_0^T (c(t)'u(t) + g(t)'y(t))\ dt + \int_0^T a(t)'\ d\pi(t) + \int_0^T b(t)'\eta(t)\ dt + \int_0^T h(t)'\ d\xi(t) \\
= & \int_0^T (c(t)'u(t) + g(t)'y(t))\ dt + \int_0^T \left(\int_0^t Gu(s)\ ds + Ey(t)\right)\ d\pi(t) + \\
& \int_0^T (Hu(t) + z(t))'\eta(t)\ dt + \int_0^T (Fy(t) + \bar{z}(t))'d\xi(t) \\
= & \int_0^T (c(t)'u(t) + g(t)'y(t))\ dt - \int_0^T \pi(t)'Gu(t)\ dt + \int_0^T Ey(t)\ d\pi(t) + \\
& \int_0^T (Hu(t) + z(t))'\eta(t)\ dt + \int_0^T (Fy(t) + \bar{z}(t))'d\xi(t) \\
= & \int_0^T (c(t) - G'\pi(t) + H'\eta(t))'\, u(t)\ dt + \\
& \int_0^T y(t)'\ d\left(E'\pi(t) + F'\xi(t) - \int_t^T g(t)\ dt\right) +
\end{aligned}
$$

$$\begin{aligned}
& \int_0^T z(t)'\eta(t)\,dt + \int_0^T \bar{z}(t)'d\xi(t) \\
= & \int_0^T (c(t) - G'\pi(t) + H'\eta(t))'\,u(t)\,dt + \int_0^T z(t)'\eta(t)\,dt + \int_0^T \bar{z}(t)'d\xi(t) \\
\geq & \ 0,
\end{aligned}$$

where $\bar{z}(t) = h(t) - Fy(t)$. □

The requirement of $\pi(t)$ being a bounded measurable VF function in $(SCSCLP^*)$ is important, it makes the integration by parts valid in the proof of the above proposition (cf. Proposition A.2 in the Appendix). As a consequence of the proof, we have the following corollary:

Corollary 2.1 *Strong duality holds between $(SCSCLP)$ and $(SCSCLP^*)$ if and only if there exist $(u(t),\ y(t))$ and $(\pi(t),\ \eta(t),\ \xi(t))$ which are feasible to $(SCSCLP)$ and $(SCSCLP^*)$ respectively, and satisfy the following conditions.*

$$\begin{aligned}
\int_0^T (c(t) - G'\pi(t) + H'\eta(t))'\,u(t)\,dt &= 0; \\
\int_0^T (b(t) - Hu(t))'\eta(t)\,dt &= 0; \\
\int_0^T (h(t) - Fy(t))'d\xi(t) &= 0. \qquad (2.1)
\end{aligned}$$

We call (2.1) the complementary slackness condition for $(SCSCLP)$ and $(SCSCLP^*)$.

2.4 Algorithms for SCLP

There are several known algorithms for $(SCLP)$.

The first type of algorithm discretizes time and transforms the problem into a very large linear programming problem. The algorithm then solves the linear programming problem by using state of the art linear programming solvers (see Buie and Abraham [17]). This approach was recently strengthened by Pullan [77] and developed further by Philpott and Craddock [74] for the network version of $(SCLP)$. Their algorithms iteratively add new breakpoints into the partition.

The second is a "simplex" type algorithm that involves the extension of concepts such as "basic solutions," "dual variables," and "pivots," see Lehman [57], Drews [29], Hartberger [45], Segers [84], Perold [72] and Anstreicher [7].

The third type of algorithm develops the primal-dual path following interior point methods in Hilbert spaces, see Ito *et. al.* [48].

By utilizing Proposition 2.3, Anderson [2] developed a convex quadratic program for $(SCLP)$ that enumerates all the extreme points of a set of linear programs.

However, all these algorithms have difficulty in solving large problems. For the first type of algorithm, the time horizon has to be discretized into so many small intervals that it takes a very long time to solve the linear program. The optimal solution usually turns out to be constant or linear on the majority of these intervals. For the second type of algorithm, to the best of our knowledge, there is no theoretical convergence guarantee and for the third type of algorithm, we need to have an initial feasible starting point that is also in the interior of the feasible region. Also, very fine discretizations are used in order to solve the inverse of some abstract operator and thus create the same difficulty as the first type of algorithm. Finally, for the fourth type of algorithm, the quadratic program proposed is of exponential size and requires all the extreme points of a set of linear programs known a priori.

As a result of this discussion, we wish to develop a new efficient algorithm for $(SCSCLP)$. We note that the computational complexity of problems $(SCLP)$ and $(SCSCLP)$ is still an open question.

Chapter 3

An Algorithm for SCSCLP

In Chapters 1 and 2, we introduced fluid models and reviewed the general results of continuous linear programming. In this chapter, we will develop a new algorithm for solving SCSCLP problems under Assumption 2.1. The new algorithm also uses discretization. Unlike other algorithms (cf. Section 2.4), it varies the discretization and control simultaneously. Based on the number of constant pieces allowed in the control, we develop a quadratic program with polyhedral constraints. The quadratic program is generally nonconvex. However, we do not need to solve the quadratic program to optimality. We only need to obtain a KKT point. We use the Frank-Wolfe method (see Martos [64] and Murty [67]) or general matrix splitting algorithms (see Lin and Pang [59], Eckstein [30], Bertsekas and Tsitsiklis [13], Luo and Tseng [62]) to find a KKT point for the quadratic program. By gradually increasing (and occasionally decreasing) the number of pieces allowed in the control, we can improve upon any nonoptimal KKT solution. We call this the Successive Quadratic Programming method. By a KKT solution structural result of Luo and Tseng [62], we show that the iterates of the algorithm move from one polyhedral set to another, with improved cost. By bounding the size of the quadratic programs we encounter, we bound the number of all such polyhedral sets. We show that the new algorithm converges in finite time. The absence of a duality gap and the existence of certain highly structured optimal solutions for ($SCSCLP$) and ($SCSCLP^*$) follow as byproducts. We will give some computational results of the new algorithm later in Chapter 6.

The remainder of this chapter is structured as follows. In Section 3.1, we reiterate problem ($SCSCLP$) and our assumptions, we then introduce some standard definitions and notations. In Section 3.2, we develop a quadratic program with polyhedral constraints. In Section 3.3, we review some nonlinear programming techniques for calculating a KKT point of a quadratic program with polyhedral constraints. In Section 3.4, we develop a procedure for removing redundant intervals in a feasible solution for ($SCSCLP$). In Section 3.5, we introduce a new discrete approximation for ($SCSCLP$) which is closely related to the dual problem. From this discrete approximation, we derive a criterion to detect whether a feasible solution is optimal for ($SCSCLP$). If the criterion is not satisfied, we point out a descent direction for the feasible solution to ($SCSCLP$). In Section 3.6, we formally state the new algorithm. In Section 3.7, we prove the finite convergence result for the new algorithm and in Section 3.8, we use the new algorithm to prove new duality results and new optimal solution structural results for ($SCSCLP$). The reader is advised to first read Sections 3.1 and 3.6, to have a general idea on the problem we are solving, the assumptions and the new algorithm, and then come back to Sections 3.1, 3.2 and so on.

3.1 Some Definitions and Notations

First, we reiterate problem ($SCSCLP$) and our assumptions. We consider the problem

$$
\begin{array}{lll}
(SCSCLP) & \textit{minimize} & \int_0^T (c(t)'u(t) + g(t)'y(t)) \; dt \\
 & \textit{subject to} & \int_0^t Gu(t) \; dt + Ey(t) = a(t) \\
 & & Hu(t) \leq b(t) \\
 & & Fy(t) \leq h(t) \\
 & & u(t) \geq 0, \qquad t \in [0, \; T]
\end{array}
$$

and its dual

$$
\begin{array}{lll}
(SCSCLP^*) & \textit{maximize} & -\int_0^T a(t)' \; d\pi(t) - \int_0^T b(t)'\eta(t) \; dt - \int_0^T h(t)' \; d\xi(t) \\
 & \textit{subject to} & c(t) - G'\pi(t) \; dt + H'\eta(t) \geq 0
\end{array}
$$

$$E'\pi(t) + F'\xi(t) = \int_t^T g(t)\, dt$$
$\pi(t)$ is a bounded measurable VF function
$\xi(t)$ monotonic increasing and right continuous
on $[0,\ T]$ with $\xi(T) = 0, \quad \pi(T) = 0$
$\eta(t) \geq 0, \qquad \text{for } t \in [0,\ T],$

under the following assumptions:

Assumption 3.1

a) $a(t)$ and $h(t)$ are continuous,

b) $a(t)$, $c(t)$ and $h(t)$ are piecewise linear,

c) $b(t)$ and $g(t)$ are piecewise constant,

d) problem $(SCSCLP)$ is feasible and its objective value is bounded from below.

We require that $u(t)$, $y(t)$ and $\eta(t)$ are bounded measurable on $[0,\ T]$.

The following are standard definitions and notations which we will use throughout the remainder of this chapter.

We call a sequence of time epochs $P = \{t_0, \cdots, t_p\}$ a partition of $[0,\ T]$ if

$$0 = t_0 \leq t_1 \leq \cdots \leq t_p = T.$$

We use $|P|$ to denote the cardinality of P. Note, since our development sometimes treats t_i as a variable, we allow $t_i = t_{i-1}$ for some $i \geq 1$ and always treat t_i and t_{i-1} as two different variables.

We say that a function $f(t)$ is piecewise constant (linear) with a partition $P = \{t_0, \cdots, t_p\}$, if $f(t)$ is constant (linear) on $[t_{i-1},\ t_i)$ for $i = 1, \cdots, p$. We say $f(t)$ is piecewise constant (linear) on $[0,\ T]$ if $f(t)$ is piecewise constant (linear) with some partition of $[0,\ T]$.

Let $P = \{t_0, \cdots, t_p\}$ be a partition of $[0,\ T]$. Throughout this chapter, we assume Assumption 2.1 holds. We also assume $a(t)$, $h(t)$ and $c(t)$ are piecewise linear and $b(t)$ and $g(t)$ are piecewise constant with partition P. Let $\mathcal{B}$ be the set of breakpoints of $a(t)$, $b(t)$, $c(t)$, $g(t)$ and $h(t)$. For each breakpoint in $\mathcal{B}$, we select one element t_i in P, such that its value denotes the same time in $[0,\ T]$ as the breakpoint. We always select t_0 for 0 and t_p for T. We denote D^P as the set of selected elements of P excluding t_0 and t_p. Let $D_1^P = D^P \bigcup \{t_0, t_p\}$. We sometimes omit the superscripts P when the context is clear.

We say an interval $[t_{i-1},\ t_i]$ is a subinterval of $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P, if $l \le i-1 < i \le m$. In this case, we also say that t_{i-1}, t_i and $[t_{i-1},\ t_i]$ reside on $[t_l,\ t_m]$.

For a function $f(t)$, we will use notations

$$f(t-) = \lim_{s\to t-} f(s) \qquad \text{and} \qquad f(t+) = \lim_{s\to t+} f(s),$$

when the above limits exist and t is not equal to any breakpoint in D_1^P. If $[t_{i-1},\ t_i]$ is a zero length subinterval of $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P. By convention, we let

$$f(t_i-) = \begin{cases} \lim_{s\to t_i-} f(s), & \text{if } t_i = t_m \\ \lim_{s\to t_i+} f(s), & \text{if } t_i = t_l, \end{cases}$$

and let $f(t_{i-1}+) = f(t) = f(t_i-)$. We note that the value of $f(t_i)$ is sensitive to both the value of t_i and its index i.

If $t_i \neq t_{i-1}$ for all i, for a set of $2p$ variables $\hat{f}(t_0+), \hat{f}(t_1-),\ \hat{f}(t_1+),\ \cdots,\ \hat{f}(t_{p-1}+), \hat{f}(t_p-)$, the function $f(t)$ defined by

$$f(t) = \begin{cases} \hat{f}(t_i+), & \text{if } t = t_0, t_1, \cdots,\ t_{p-1}, \\ 0, & \text{if } t = T, \\ \frac{t_i - t}{t_i - t_{i-1}}\hat{f}(t_{i-1}+) + \frac{t - t_{i-1}}{t_i - t_{i-1}}\hat{f}(t_i-), & \text{for } t \in (t_{i-1},\ t_i),\ i = 1, \cdots, p, \end{cases}$$

is called the piecewise linear extension of these $2p$ variables; for a set of variables $\hat{f}(t_0+),\ \hat{f}(t_1+),\ \cdots,\ \hat{f}(t_{p-1}+)$, the function $f(t)$ defined by

$$f(t) = \begin{cases} \hat{f}(t_{p-1}+), & t = T, \\ \hat{f}(t_{i-1}+), & \text{for } t \in [t_{i-1},\ t_i), \text{ for } i = 1, \cdots, p, \end{cases}$$

is called the piecewise constant extension of these p variables.

For two functions $f(t)$ and $g(t)$, we denote $\int_a^b f(t)\, dg(t)$ as the Lebesgue-Stieltjes integral of $f(t)$ with respect to $g(t)$ from a to b, given that the integral exists, including both a and b. For any mathematical program (LP) we shall write $V(\text{LP})$ for the optimal value of the objective function, which may not be attained. For any n dimensional vector x, we denote by x_i the i-th coordinate of x, and, for any nonempty subset $Q \subseteq \{1, \cdots, n\}$, we use x_Q, $[x]_Q$ or $(x)_Q$ to denote the the vector with components x_i, $i \in Q$ (with x_i arranged in the same order as in x). For a matrix A, we denote by A_{ij} the j-th element of the i-th row of matrix A and denote by $A_{i\bullet}$ the i-th row of A.

3.2 A Quadratic Programming Sub-problem

By Proposition 2.3, there exists an optimal basic feasible solution to $(SCLP)$ whose $u(t)$ is piecewise constant (see Figure 3-1) when Assumption 3.1 holds and the solution set to $(SCLP)$ is bounded. We will prove later in the chapter that this remains true for $(SCSCLP)$. For any feasible control $u(t)$ that is piecewise constant with respect to partition P, we have the following standard linear approximation problem (see Pullan [77] and the references therein).

$$
\begin{array}{lll}
DP(P) & \min & \sum_{i=1}^{p}(t_i - t_{i-1})\hat{u}(t_{i-1}+)'c(\frac{t_i + t_{i-1}}{2}) + \sum_{i=1}^{p} \frac{t_i - t_{i-1}}{2}(\hat{y}(t_i) + \hat{y}(t_{i-1}))'g(t_{i-1}+) \\
& s.t. & E\hat{y}(t_0) = a(t_0), \\
& & (t_i - t_{i-1})G\hat{u}(t_{i-1}+) + E\hat{y}(t_i) - E\hat{y}(t_{i-1}) = a(t_i) - a(t_{i-1}), \\
& & \quad i = 1, \cdots, p, \\
& & H\hat{u}(t_{i-1}+) \leq b(t_{i-1}+), \qquad i = 1, \cdots, p, \\
& & F\hat{y}(t_i) \leq h(t_i), \qquad i = 0, \cdots, p, \\
& & \hat{u}(t_{i-1}+) \geq 0 \qquad i = 1, \cdots, p,
\end{array}
$$

where we have the convention that $c(\frac{t_i+t_{i-1}}{2}) = c(t_i-)$, whenever $t_i = t_{i-1}$. Note that even though it is possible that $t_i = t_{i-1}$, for some $i \geq 1$, we still treat $\hat{u}(t_i+)$ and $\hat{u}(t_{i-1}+)$ as separate variables. If $(\hat{u},\ \hat{y})$ is a feasible solution to $DP(P)$, where partition P satisfies $t_i \neq t_{i-1}$ for all i, the piecewise constant extension of $\hat{u}$ together

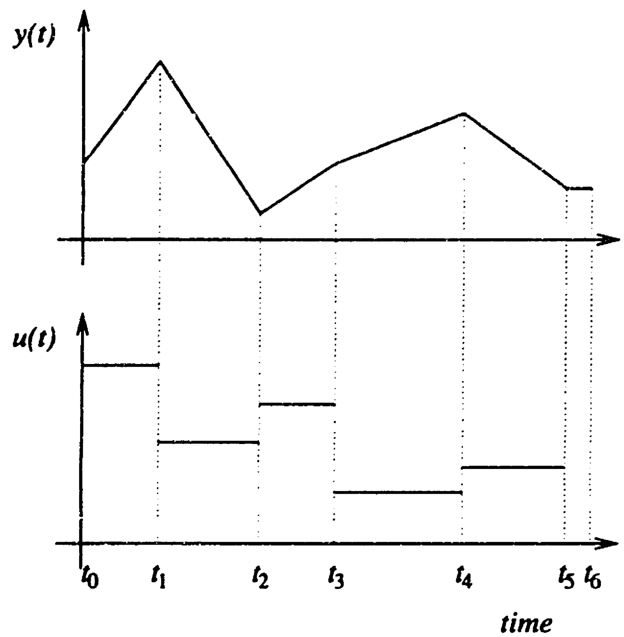

Figure 3-1: A piecewise constant optimal control for $(SCLP)$

with the piecewise linear extension of $\hat{y}$, defines a feasible solution to $(SCSCLP)$ with the same cost, due to Assumption 2.1. If we fix the partition, $DP(P)$ is a linear programming problem. So, once an optimal partition P is known, an optimal solution can be computed by solving the linear program $DP(P)$.

However, we do not know the optimal partition in advance. The algorithms proposed by Pullan [77] and by Philpott and Craddock [74] alternatively do the following two steps:

1) Improve the control for the current partition.

2) Improve the partition.

In contrast, the algorithm we propose improves both the control and partition at the same time. By introducing new variables

$$\hat{v}(t_i) = (t_i - t_{i-1})\hat{u}(t_{i-1}+), \tag{3.1}$$

we can eliminate variables $\hat{u}$ and obtain the following simpler mathematical programming problem in variables $\hat{v}$, $\hat{y}$ and $\hat{t}$, with $\hat{t}$ being the vector of t_is such that $t_i \notin D_1^P$.

$$\begin{aligned}
QP(|P|) \quad \min \quad & \sum_{i=1}^{p} \hat{v}(t_i)'c(\frac{t_i + t_{i-1}}{2}) + \sum_{i=1}^{p} \frac{t_i - t_{i-1}}{2}(\hat{y}(t_i) + \hat{y}(t_{i-1}))'g(t_{i-1}+) \\
s.t. \quad & E\hat{y}(t_0) = a(t_0), \\
& G\hat{v}(t_i) + E\hat{y}(t_i) - E\hat{y}(t_{i-1}) = a(t_i) - a(t_{i-1}), \qquad (3.2) \\
& \qquad i = 1, \cdots, p, \\
& H\hat{v}(t_i) \leq (t_i - t_{i-1})b(t_{i-1}+), \qquad i = 1, \cdots, p, \\
& F\hat{y}(t_i) \leq h(t_i) \qquad i = 0, \cdots, p, \\
& 0 = t_0 \leq t_1 \leq \cdots \leq t_p = T, \\
& \hat{v}(t_i) \geq 0 \qquad i = 1, \cdots, p,
\end{aligned}$$

where $c(\frac{t_i+t_{i-1}}{2}) = c(t_i-)$ whenever $t_i = t_{i-1}$. Note that the breakpoints in D_1^P are fixed and are not variables. We treat both $\hat{v}(t_i)$ and $\hat{y}(t_i)$ as variables. Let t_l and t_m be two consecutive breakpoints in D_1^P. For any $i \in (l,\ m]$, $c(\frac{t_i+t_{i-1}}{2})$, $a(t_i) - a(t_{i-1})$ and $h(t_i)$ are the following linear functions of t_i and t_{i-1}:

$$\begin{aligned}
c(\frac{t_i + t_{i-1}}{2}) &= c(t_l) + \frac{t_i + t_{i-1} - 2t_l}{2}\dot{c}(t_l+) \\
a(t_i) - a(t_{i-1}) &= (t_i - t_{i-1})\dot{a}(t_l+) \\
h(t_i) &= h(t_l) + (t_i - t_l)\dot{h}(t_l+),
\end{aligned}$$

and $g(t_{i-1}+) = g(t_l)$ and $b(t_{i-1}+) = b(t_l)$ are constant vectors. So, $QP(|P|)$ is a quadratic programming problem with polyhedral constraints.

Given a feasible solution $(\hat{v},\ \hat{y},\ \hat{t})$ to $QP(|P|)$ such that $t_i \neq t_{i-1}$ for all i, we can obtain a feasible solution $(\hat{u}, \hat{y})$ to problem $DP(P)$ with P defined from vector $\hat{t}$ and the breakpoints in D_1^P and $\hat{u}$ defined from

$$\hat{u}(t_{i-1}+) \;=\; \frac{\hat{v}(t_i)}{t_i - t_{i-1}}. \qquad (3.3)$$

Equation (3.1) defines an injective mapping from the solution set to $DP(P)$ to the solution set to $QP(|P|)$. The two related solutions have the same solution value.

However, if $t_i = t_{i-1}$ but $\hat{v}(t_i) \neq 0$ for some i, the right hand side of (3.3) is not properly defined, i.e., there may be a solution to $QP(|P|)$ for which the corresponding solution to $DP(P)$ can not be constructed. We overcome this difficulty by constantly removing redundant zero length intervals in a feasible solution and by using only the solution $(\hat{v},\ \hat{y},\ \hat{t})$ to $QP(|P|)$ that satisfies

$$t_i \neq t_{i-1} \quad \text{for all} \quad i \geq 1 \tag{3.4}$$

to construct a feasible solution for $DP(P)$ (and so for $(SCSCLP)$). When some zero length intervals can not be removed, we show there is a series of feasible solutions to $QP(|P|)$ that satisfies (3.4) whose solution value becomes arbitrarily close to that of the feasible solution to $QP(|P|)$. This is key to understanding the absence of a duality gap result between $(SCSCLP)$ and $(SCSCLP^*)$, as we will see later on.

Lemma 3.1 *Suppose $u(t)$ in the feasible solutions to $(SCSCLP)$ is bounded. Let $(\hat{v},\ \hat{y},\ \hat{t})$ be a feasible solution to $QP(|P|)$, the following result holds.*

$$\hat{v}(t_i) = 0 \quad \text{whenever} \quad t_i = t_{i-1}. \tag{3.5}$$

Proof Suppose $t_i = t_{i-1}$ for some i, but $\hat{v}(t_i) \neq 0$. Let $[t_l,\ t_m]$ be the interval t_i resides on, where t_l and t_m are two consecutive breakpoints in D_1^P. Without loss of generality, we may assume there exists a positive length subinterval of $[t_l,\ t_m]$ that is adjacent to $[t_{i-1},\ t_i]$ (since we can switch the values of $\hat{v}(t_j)$ and $\hat{y}(t_j)$ with adjacent zero length subintervals on $[t_l,\ t_m]$ and maintain the feasibility of the solution). We assume the adjacent positive length subinterval on $[t_l,\ t_m]$ is $[t_{i-2},\ t_{i-1}]$. When the adjacent positive length subinterval of $[t_l,\ t_m]$ is $[t_i,\ t_{i+1}]$, a similar analysis applies.

For any $\tau \in (0,\ 1)$, it is easy to verify the following solution is feasible for $QP(|P|)$:

$$\begin{aligned}
\tilde{t}_j^{\tau} &= \begin{cases} t_j; & \text{if } j \neq i-1 \text{ and } j \neq i, \\ \tau t_{i-2} + (1-\tau)t_{i-1}; & \text{if } j = i-1, \\ t_i; & \text{if } j = i, \end{cases} \\
\tilde{v}^{\tau}(t_j^{\tau}) &= \begin{cases} \hat{v}(t_j); & \text{if } j \neq i-1 \text{ and } j \neq i, \\ (1-\tau)\hat{v}(t_{i-1}); & \text{if } j = i-1, \\ \tau\hat{v}(t_{i-1}) + \hat{v}(t_i); & \text{if } j = i, \end{cases}
\end{aligned}$$

$$\tilde{y}^{\tau}(t_j^{\tau}) = \begin{cases} \hat{y}(t_j); & \text{if } j \neq i-1 \text{ and } j \neq i, \\ \tau\hat{y}(t_{i-2}) + (1-\tau)\hat{y}(t_{i-1}); & \text{if } j = i-1, \\ \hat{y}(t_i); & \text{if } j = i. \end{cases}$$

The basic idea is to split interval $[t_{i-2},\ t_{i-1}]$ into two intervals $[t_{i-2},\ \tau t_{i-2}+(1-\tau)t_{i-1}]$ and $[\tau t_{i-2} + (1-\tau)t_{i-1},\ t_{i-1}]$, and combine the second interval with $[t_{i-1},\ t_i]$. It is easy to check that $(\tilde{v}^{\tau},\ \tilde{y}^{\tau},\ \tilde{t}^{\tau})$ is feasible for $QP(|P|)$ and has one less zero length interval than $(\hat{v},\ \hat{y},\ \hat{t})$. Applying the same process repeatedly, we can eliminate all the zero length intervals in solution $(\hat{v},\ \hat{y},\ \hat{t})$.

Let $(\bar{v}^{\tau},\ \bar{y}^{\tau},\ \bar{t}^{\tau})$ be the resulting solution and let Q be the resulting partition. Hence $(\bar{v}^{\tau},\ \bar{y}^{\tau},\ \bar{t}^{\tau})$ is feasible for $QP(|Q|)$. From this solution, we can construct a feasible solution for $DP(Q)$ (and thus for $(SCSCLP)$) by using (3.3). However, as τ tends to zero, the corresponding feasible solution to $DP(P)$ is unbounded from above (since the denominator in (3.3) goes to zero but the numerator is bounded away from zero). Thus $u(t)$ in $(SCSCLP)$ is unbounded and this creates a contradiction.

□

We remark that Lemma 3.1 implies that if $u(t)$ is bounded and E is an identity matrix (e.g., a bounded and feasible $(SCLP)$), then $\hat{y}(t_{i-1}) = \hat{y}(t_i)$ whenever $t_{i-1} = t_i$. In general, when (3.5) holds, it is possible that $\hat{y}(t_{i-1}) \neq \hat{y}(t_i)$ even if $t_{i-1} = t_i$. If in addition to (3.5), $\hat{y}(t_{i-1}) = \hat{y}(t_i)$ for some i such that $t_{i-1} = t_i$, then we can eliminate the zero length interval $[t_{i-1},\ t_i]$ from $(\hat{v},\ \hat{y},\ \hat{t})$ while maintaining the feasibility and solution value of the solution. This fact will be used later in Section 3.4 to remove redundant intervals.

In general, $u(t)$ may not be bounded in a feasible solution to $(SCSCLP)$. It is possible that there is no feasible solution to $(SCSCLP)$ that is optimal for $(SCSCLP)$. This perhaps is the key difficulty in establishing the absence of a duality gap between $(SCSCLP)$ and $(SCSCLP^*)$ by conventional methods. Hence, we have the following relationship between $(SCSCLP)$ and $QP(|P|)$.

Lemma 3.2 *Given any feasible solution $(\hat{v},\ \hat{y},\ \hat{t})$ to $QP(|P|)$, there exists a series of feasible solutions $(\hat{v}^k,\ \hat{y}^k,\ \hat{t}^k)$ to $QP(|P|)$ that satisfies* (3.4) *and whose solution value becomes arbitrarily close to that of $(\hat{v},\ \hat{y},\ \hat{t})$ as k tends to infinity.*

Proof By using the same procedure used to prove Lemma 3.1, we can construct a solution $(\bar{v}^{\tau},\ \bar{y}^{\tau},\ \bar{t}^{\tau})$ which is feasible to $QP(|Q|)$ for some partition Q and satisfies

(3.4). It is easily verified that the solution value of $(\bar{v}^\tau,\ \bar{y}^\tau,\ \bar{t}^\tau)$ to $QP(|Q|)$ becomes arbitrarily close to that of $(\hat{v},\ \hat{y},\ \hat{t})$ as τ goes to zero. □

In fact, we can have $t_i \neq t_{i+1}$ and $t_{i-1} \neq t_{i-2}$ whenever $t_i = t_{i-1}$ in a local optimum for $QP(|P|)$. The existence of $\hat{v}(t_i) \neq 0$ but $t_i = t_{i-1}$ indicates the presence of the Dirac δ function in $u(t)$ at time t_i.

A direct consequence of Lemma 3.2 is $V((SCSCLP)) \leq V(QP(|P|))$ for all P. This fact enables us to solve $(SCSCLP)$ through solving $QP(|P|)$ for a series of partitions. We note that $V(QP(|P|)) = V(SCSCLP)$ does not imply there is a feasible solution for $(SCSCLP)$, whose solution value is equal to $V(QP|P|)$, due to the possible presence of zero length intervals in P.

3.3 Finding a KKT Point for $QP(|P|)$

We do not need to solve the nonconvex quadratic program $QP(|P|)$ to optimality, as we will see in Section 3.6. We only need to compute a series of KKT points (or equivalently, stationary points) of a set of quadratic programs. In this section, we examine several nonlinear programming techniques for finding a KKT point of a quadratic program.

3.3.1 The Frank-Wolfe Method

There is a well developed algorithm called the Frank-Wolfe method (otherwise known as the *conditional gradient method*) for solving quadratic programming problems with polyhedral constraints. It is often used to calculate a KKT point for the following problem:

$$\begin{array}{lll} (NLP) & minimize & \theta(x) \\ & subject\ to & x \in X, \end{array}$$

where X is the following polyhedral set

$$X = \{\ x \mid A_{i\bullet}x = b_i \ \text{ for } i = 1 \text{ to } m; A_{i\bullet}x \geq b_i \ \text{ for } i = m+1 \text{ to } m+p\} \qquad (3.6)$$

and $\theta(\cdot)$ is the following quadratic function:

$$\theta(x) = \frac{1}{2}x'Mx + q'x,$$

and M is a symmetric matrix.

Given a feasible solution $x \in X$, the first order necessary conditions for x to be locally optimal for problem (NLP) are that there exists $\pi = (\pi_1, \cdots, \pi_{m+p})'$ such that

$$\begin{aligned} \nabla\theta(x)' &= \sum_{i=1}^{m+p} \pi_i A_{i\bullet} \\ \pi_i &\geq 0 \quad \text{for } i = m+1 \text{ to } m+p \\ \pi_i(A_{i\bullet}x - b_i) &= 0 \quad \text{for } i = m+1 \text{ to } m+p. \end{aligned}$$

We call $x \in X$ a KKT point of the system (NLP) if it satisfies the above conditions.

The Frank-Wolfe method starts with an initial feasible solution x^0 and generates a sequence of feasible points $\{ x^k \ : k = 1, 2, \cdots, \}$ satisfying $\theta(x^{k+1}) < \theta(x^k)$ for all k.

The Frank-Wolfe Method

At the k–th iteration we are given an $x^k \in X$ ($x^0 \in X$ is given). Solve the following linear program in the variable y^k:

$$\begin{aligned} (SUBLP^k) \quad & \textit{minimize} \quad (\nabla\theta(x^k))'y^k \\ & \textit{subject to} \quad x^k + y^k \in X. \end{aligned}$$

If $(\nabla\theta(x^k))'y^k = 0$, the algorithm terminates; otherwise the algorithm does a line search to find the minimum of $\theta(x^k + \alpha y^k)$ subject to $0 \leq \alpha \leq 1$. Let α^k be the minimum of this line search problem. Let $x^{k+1} = x^k + \alpha^k y^k$ and the algorithm continues to the next iteration.

Note that 0 is a trivial feasible solution for $(SUBLP^k)$. Let y^k be the optimal solution of $(SUBLP^k)$ and so $(\nabla\theta(x^k))'y^k \leq 0$. If $(\nabla\theta(x^k))'y^k = 0$, 0 is optimal for $(SUBLP^k)$. Let π be the dual optimal solution for $(SUBLP^k)$. The complementary slackness condition between $(SUBLP^k)$ and its dual translates exactly to the first order necessary optimality condition of x^k and we terminate the algorithm. Otherwise, we have $(\nabla\theta(x^k))'y^k < 0$ and y^k is a descent direction at x^k. The line search guarantees $\theta(x^{k+1}) < \theta(x^k)$ for all k.

The following convergence result for this method can be found in Murty [67] (see also Martos [64]).

Theorem 3.1 *Suppose $X \neq \emptyset$. If the method does not terminate after a finite number of steps and the sequence $\{ x^k : k = 1, 2, \cdots, \}$ generated by the above method has at least one limit point, then every limit point of this sequence is a KKT point of (NLP).*

We note that in order to solve $(SUBLP^k)$, we can use either the simplex method or an interior point algorithm to compute y^k. Our computational experiments indicate that the convergence rate for the Frank-Wolfe method might be sublinear and thus slow. Another drawback of directly applying the Frank-Wolfe method is that the iterates it generates may have no limit point. For this reason, we can use another more general and more sophisticated method called the matrix splitting algorithm (see Luo and Tseng [62] and the references therein).

3.3.2 The Matrix Splitting Algorithm

The following is a matrix splitting algorithm for symmetric affine variational inequality (or AVI in short) problems. A symmetric AVI problem is to find an $x^* \in X$ satisfying

$$(Mx^* + q)'(x - x^*) \geq 0 \qquad \forall x \in X.$$

By replacing $x - x^*$ in the above problem by y and following the arguments after the problem $(SUBLP^k)$, we see that x^* is a solution to the AVI problem if and only if it is a KKT point for (NLP).

Let (B, C) be a splitting of M, i.e.,

$$M = B + C. \tag{3.7}$$

The Matrix Splitting Algorithm for solving the AVI problem, based on splitting (B, C) is the following:

The Matrix Splitting Algorithm
At the k–th iteration we are given an $x^k \in X$ ($x^0 \in X$ is given), and we compute a new iterate x^{k+1} in X satisfying

$$x^{k+1} = [x^{k+1} - Bx^{k+1} - Cx^k - q]^+, \tag{3.8}$$

where $[\cdot]^+$ is the projection mapping onto X.

Under the assumption that B and $B - C$ are positive definite and $\theta(x)$ is bounded from below on X, the following holds (see also Theorem 3.1 of Luo and Tseng [62]).

Theorem 3.2 *If B and $B - C$ are positive definite and $\theta(x)$ is bounded from below on X, then the iterates generated by the Matrix Splitting Algorithm would converge at least linearly to a KKT point to (NLP).*

Note that when B is symmetric positive semidefinite, (3.8) can be solved as a convex quadratic programming problem, see Bertsekas and Tsitsiklis [13]. The same as in the Frank-Wolfe method, the solution value in (NLP) for the iterates generated by the Matrix Splitting Algorithm decreases monotonically. Indeed, other methods for obtaining a KKT point for (NLP) exist, such as those proposed by Ye [100] and Kojima *et. al.* [55]. The solution values for the iterates of these two methods, however, do not decrease monotonically.

3.4 Removing Redundant Intervals

After finding a KKT point of $QP(|P|)$, it is possible that some zero length intervals can be removed, as we noted following Lemma 3.1. It is also possible that some adjacent intervals can be merged while maintaining the solution value. The reduction of unnecessary control pieces in the solution is a key feature of the new algorithm. This enables us to prove the convergence of the new algorithm without requiring the norm of the maximal length interval in the discretization to tend to zero (cf. Pullan [77] and Philpott and Craddock [74]).

To do this, let $(\hat{v},\ \hat{y},\ \hat{t})$ be a feasible solution to $QP(|P|)$ and, let $[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$ be two adjacent intervals that reside on $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P. We eliminate t_i from P (or equivalently, combine

$[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$) and define a new feasible solution $(\tilde{v},\ \tilde{y},\ \tilde{t})$ for $QP(|P \setminus \{t_i\}|)$ as follows. Let $\tilde{v}$ be the vector formed by removing $\hat{v}(t_{i+1})$ from $\hat{v}$ and then replacing $\hat{v}(t_i)$ with $\hat{v}(t_i) + \hat{v}(t_{i+1})$; let $\tilde{y}$ be the vector formed by removing $\hat{y}(t_i)$ from $\hat{y}$ and let $\tilde{t}$ be the vector formed by removing t_i from $\hat{t}$.

Lemma 3.3 *Let $[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$ be two adjacent intervals that reside on $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P. If*

$$\begin{aligned} & (t_{i+1} - t_i)\hat{v}(t_i)'\dot{c}(t_{i-1}+) + (t_{i+1} - t_i)\hat{y}(t_{i-1})'g(t_{i-1}+) + (t_i - t_{i-1})\hat{y}(t_{i+1})'g(t_{i-1}+) \\ \leq\ & (t_i - t_{i-1})\hat{v}(t_{i+1})'\dot{c}(t_{i-1}+) + (t_{i+1} - t_{i-1})\hat{y}(t_i)'g(t_{i-1}+), \end{aligned} \tag{3.9}$$

then we can combine $[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$ while maintaining the feasibility and the solution value of a feasible solution to $QP(|P|)$.

Proof Consider the difference between the solution value of $(\hat{v},\ \hat{y},\ \hat{t})$ and that of the solution with t_i removed from the partition P. After some simplification, we see that the new solution has a smaller solution value if and only if (3.9) holds. □

A direct corollary to Lemma 3.3 is the following.

Corollary 3.1 *Let t_l and t_m be two consecutive breakpoints in D_1^P. We can combine adjacent zero length intervals in $[t_l,\ t_m]$ and maintain the feasibility and the solution value of a feasible solution to $QP(|P|)$.*

Proof Let $[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$ be two adjacent zero length intervals that reside on $[t_l,\ t_m]$. Since $t_{i-1} = t_i = t_{i+1}$, (3.9) is trivially satisfied. By Lemma 3.3, we can combine $[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$ and maintain the feasibility and solution value of the feasible solution to $QP(|P|)$. □

By Corollary 3.1, we can combine adjacent zero length intervals. The following lemma implies that all the zero length intervals except those at the breakpoints in D_1^P can be eliminated.

Lemma 3.4 *Let $[t_{i-1},\ t_i]$ be a zero length interval that resides on $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P. Suppose $[t_{i-2},\ t_{i-1}]$ and $[t_i,\ t_{i+1}]$ are two positive length intervals that also reside on $[t_l,\ t_m]$. We can either*

1. *combine $[t_{i-2},\ t_{i-1}]$ and $[t_{i-1},\ t_i]$ or*

2. combine $[t_{i-1},\ t_i]$ *and* $[t_i,\ t_{i+1}]$

while maintaining thc feasibility and the solution value of the feasible solution to $QP(|P|)$.

Proof Since $t_{i-1} = t_i$, by Lemma 3.3, we can combine $[t_{i-2},\ t_{i-1}]$ and $[t_{i-1},\ t_i]$ if the following relation holds:

$$(t_{i-1} - t_{i-2})\hat{y}(t_i)'g(t_{i-2}+) \leq (t_{i-1} - t_{i-2})(\hat{v}(t_i)'\dot{c}(t_{i-2}+) + \hat{y}(t_{i-1})'g(t_{i-2}+)). \quad (3.10)$$

By Lemma 3.3 again, we can combine $[t_{i-1},\ t_i]$ and $[t_i,\ t_{i+1}]$ if the following relation holds:

$$(t_{i+1} - t_i)\hat{y}(t_i)'g(t_i+) \geq (t_{i+1} - t_i)(\hat{v}(t_i)'\dot{c}(t_i+) + \hat{y}(t_{i-1})'g(t_i+)). \quad (3.11)$$

By assumption, we have $t_{i+1} - t_i > 0$ and $t_{i-1} - t_{i-2} > 0$. By the linearity of $c(t)$ and the constancy of $g(t)$ on $[t_l,\ t_m]$, we have

$$g(t_{i-2}+) = g(t_i+), \quad \text{and} \quad \dot{c}(t_{i-2}+) = \dot{c}(t_i+).$$

So either (3.10) or (3.11) is true. This proves the lemma. □

Now we propose the following procedure for removing redundant intervals on $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P.

Procedure *PURIFY*

1. Repeatedly combine two adjacent intervals $[t_{i-1}, t_i]$ and $[t_i, t_{i+1}]$ in $[t_l,\ t_m]$ if (3.9) is satisfied.

When more than one pair of adjacent intervals satisfy (3.9), we can combine them in an arbitrary order, one pair at a time. Let $\tilde{P}$ be the resulting partition of $[0,\ T]$ after we apply the above procedure to P for all consecutive breakpoints in D_1^P. We call $\tilde{P}$ a *purified partition* of $[0,\ T]$. Note that the remaining zero length intervals are located at the breakpoints in D_1^P and there are at most $2|D_1^P|$ zero length intervals in P.

3.5 Improving a Nonoptimal Solution

One major step of the new algorithm is to calculate a KKT point of the system $QP(|P|)$ for some partition P of $[0,\ T]$. However, the problem $QP(|P|)$ is usually nonconvex. To obtain a global optimal solution for $(SCSCLP)$, we must be able to improve a solution that is not globally optimal for $(SCSCLP)$. In this section, we give descent directions for the solution that is not globally optimal for $(SCSCLP)$. To do so, we first introduce a new discrete approximation for $(SCSCLP)$ which is closely related to the dual problem $(SCSCLP^*)$. From this new approximation, we derive a criterion that detects whether a solution is globally optimal for $(SCSCLP)$. If this criterion is not satisfied, we give a descent direction for the current solution and thus improve the solution value. We show that instead of using the direction constructed in Section 3.5.3, an algorithm for $(SCSCLP)$ can also use the Frank-Wolfe method or the Matrix Splitting Algorithm to find a descent direction. We also show the first iterate of the Frank-Wolfe method provides an upper bound on the current duality gap.

3.5.1 A New Discrete Approximation

For partition $P = \{t_0, \cdots, t_p\}$, we let $P' = \left\{t_0, \frac{t_0+t_1}{2}, t_1, \cdots, \frac{t_{p-1}+t_p}{2}, t_p\right\}$ be a refined partition of P. Consider the following new discrete approximation to $(SCSCLP)$, a close variation of the second discretization in Pullan [77]:

$$
\begin{aligned}
AP1(P)\quad \min\quad & \sum_{i=1}^{p} \frac{t_i - t_{i-1}}{2}\left(c(t_{i-1}+)'\hat{u}(t_{i-1}+) + c(t_i-)'\hat{u}(t_i-) + 2\hat{y}\left(\frac{t_{i-1}+t_i}{2}\right)' g(t_{i-1}+)\right) \\
s.t.\quad & E\hat{y}(t_0) = a(t_0), \\
& \left(\frac{t_i - t_{i-1}}{2}\right) G\hat{u}(t_i-) + E\hat{y}(t_i) - E\hat{y}\left(\frac{t_i+t_{i-1}}{2}\right) = a(t_i) - a\left(\frac{t_i+t_{i-1}}{2}\right), \\
& \qquad i = 1, \cdots, p, \\
& \left(\frac{t_i - t_{i-1}}{2}\right) G\hat{u}(t_{i-1}+) + E\hat{y}\left(\frac{t_i+t_{i-1}}{2}\right) - E\hat{y}(t_{i-1}) = a\left(\frac{t_i+t_{i-1}}{2}\right) - a(t_{i-1}), \\
& \qquad i = 1, \cdots, p, \\
& H\hat{u}(t_{i-1}+) \le b(t_{i-1}+), \qquad i = 1, \cdots, p, \\
& H\hat{u}(t_i-) \le b(t_i-), \qquad i = 1, \cdots, p, \\
& F\hat{y}(t_i) \le h(t_i), \qquad i = 0, \cdots, p,
\end{aligned}
$$

$$F\hat{y}\left(\frac{t_i+t_{i-1}}{2}\right) \leq h\left(\frac{t_i+t_{i-1}}{2}\right), \qquad i=1,\cdots,p,$$
$$\hat{u}(t_i-), \hat{u}(t_{i-1}+) \geq 0, \quad i=1,\cdots,p.$$

Problem $AP1(P)$ is closely related to the dual problem. The linear programming dual of $AP1(P)$ gives rise to feasible solutions for the dual problem ($SCSCLP^*$). Thus an optimal solution to $AP1(P)$ contains the dual information. We will construct a descent solution for ($SCSCLP$) based on a solution for $AP(P)$, a closely related linear program, to be defined shortly.

It is clear that the set of feasible solutions to $AP1(P)$ is the same as the set of feasible solutions to $DP(P')$ if we identify $\hat{u}(t_i-)$ in $AP1(P)$ with $\hat{u}((\frac{t_{i-1}+t_i}{2})+)$ in $DP(P')$. There are two differences between $DP(P')$ and $AP1(P)$, both of which reside in the objective function. First, instead of averaging the cost coefficients of $u(t)$ over each subinterval, the instantaneous values of the cost coefficients at the original breakpoints of P are used. Second, instead of using the average values of the state variable $y(t)$ in each subinterval, the values of $y(t)$ at the midpoint of each subinterval of P are used. It is a fact that any feasible solution for $DP(P)$ defines a feasible solution for $DP(P')$ and thus for $AP1(P)$, and these two solutions have the same solution value.

Similar to $QP(|P|)$, we introduce $\hat{v}$ to eliminate $\hat{u}$, where

$$\hat{v}(t_{i-1}+) = \frac{t_i-t_{i-1}}{2}\hat{u}(t_{i-1}+) \quad \text{and} \quad \hat{v}(t_i-) = \frac{t_i-t_{i-1}}{2}\hat{u}(t_i-). \tag{3.12}$$

Now $AP1(P)$ is transformed into the following linear program in $\hat{v}$ and $\hat{y}$:

$$\begin{aligned}
AP(P) \quad \min \quad & \sum_{i=1}^{p}\left(c(t_{i-1}+)'\hat{v}(t_{i-1}+) + c(t_i-)'\hat{v}(t_i-) + (t_i-t_{i-1})\hat{y}\left(\frac{t_{i-1}+t_i}{2}\right)' g(t_{i-1}+)\right) \\
s.t. \quad & E\hat{y}(t_0) = a(t_0), \\
& G\hat{v}(t_i-) + E\hat{y}(t_i) - E\hat{y}\left(\frac{t_i+t_{i-1}}{2}\right) = a(t_i) - a\left(\frac{t_i+t_{i-1}}{2}\right), \\
& \qquad i=1,\cdots,p, \\
& G\hat{v}(t_{i-1}+) + E\hat{y}\left(\frac{t_i+t_{i-1}}{2}\right) - E\hat{y}(t_{i-1}) = a\left(\frac{t_i+t_{i-1}}{2}\right) - a(t_{i-1}), \\
& \qquad i=1,\cdots,p, \\
& H\hat{v}(t_{i-1}+) \leq \left(\frac{t_i-t_{i-1}}{2}\right) b(t_{i-1}+), \qquad i=1,\cdots,p,
\end{aligned}$$

$$
\begin{aligned}
&H\hat{v}(t_i-) \le \left(\frac{t_i - t_{i-1}}{2}\right) b(t_i-), \qquad i = 1, \cdots, p, \\
&F\hat{y}(t_i) \le h(t_i), \qquad i = 0, \cdots, p, \\
&F\hat{y}\left(\frac{t_i + t_{i-1}}{2}\right) \le h\left(\frac{t_i + t_{i-1}}{2}\right), \qquad i = 1, \cdots, p, \\
&\hat{v}(t_i-),\ \hat{v}(t_{i-1}+) \ge 0, \quad i = 1, \cdots, p.
\end{aligned}
$$

Similar to $AP1(P)$ and $DP(P')$, $AP(P)$ and $QP(|P'|)$ have the same feasible solution set if the partition in $QP(|P'|)$ is fixed to P'. We note that the actual value of $\hat{y}(t_0)$ does not affect the objective value of $AP(P)$ as long as $E\hat{y}(t_0) = a(t_0)$ and $F\hat{y}(t_0) \le h(t_0)$ (which is indeed feasible by assumption). The dual problem for $AP(P)$ (after eliminating $\hat{y}(t_0)$) can be written as

$$
\begin{aligned}
AP^*(P) \quad \max \quad & \hat{\pi}(t_0+)'a(t_0) \\
& + \sum_{i=1}^{p} (\hat{\pi}(t_{i-1}+) + \hat{\pi}(t_i-))' \left(a(t_i) - a\left(\frac{t_i + t_{i-1}}{2}\right)\right) \\
& - \sum_{i=1}^{p} \left(\frac{t_i - t_{i-1}}{2}\right) (\hat{\eta}(t_{i-1}+) + \hat{\eta}(t_i-))' b(t_i-) \\
& + \sum_{i=1}^{p} \left(\hat{\xi}(t_i)'h(t_i) + \hat{\xi}\left(\frac{t_{i-1} + t_i}{2}\right)' h\left(\frac{t_{i-1} + t_i}{2}\right)\right) \\
s.t. \quad & c(t_i-) - G'\hat{\pi}(t_i-) + H'\hat{\eta}(t_i-) \ge 0 \quad i = 1, \cdots, p, \\
& c(t_{i-1}+) - G'\hat{\pi}(t_{i-1}+) + H'\hat{\eta}(t_{i-1}+) \ge 0 \quad i = 1, \cdots, p, \\
& E'(-\hat{\pi}(t_i-) + \hat{\pi}(t_{i-1}+)) + F'\hat{\xi}\left(\frac{t_{i-1} + t_i}{2}\right) = (t_i - t_{i-1})g(t_{i-1}+) \\
& \qquad\qquad i = 1, \cdots, p, \\
& E'(-\hat{\pi}(t_i+) + \hat{\pi}(t_i-)) + F'\hat{\xi}(t_i) = 0 \quad i = 1, \cdots, p-1, \\
& E'(\hat{\pi}(t_p-)) + F'\hat{\xi}(t_p) = 0 \\
& \hat{\eta}(t_i-), \hat{\eta}(t_{i-1}+) \ge 0 \quad i = 1, \cdots, p, \\
& \hat{\xi}(t_i), \hat{\xi}\left(\frac{t_{i-1} + t_i}{2}\right) \le 0 \quad i = 1, \cdots, p.
\end{aligned}
$$

Similar to the second discretization in Pullan [77], the importance of $AP(P)$ lies in the fact that feasible solutions for its dual problem $AP^*(P)$ can be used either to define a feasible solution for $(SCSCLP^*)$ with the same solution value or to define a sequence of feasible solutions for $(SCSCLP^*)$ whose solution value converges to that of the original solution to $AP^*(P)$, as shown in the following theorem.

Theorem 3.3 *Suppose that P is a purified partition of $[0,\ T]$ (as defined at the end of Section 3.4). Given any feasible solution $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ to $AP^*(P)$, if (3.4) holds for P, then there exists a feasible solution $(\pi(t),\ \eta(t),\ \xi(t))$ to $(SCSCLP^*)$ whose solution value equals that of $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$. Otherwise, there exists a series of feasible solutions $(\pi^k(t),\ \eta^k(t),\ \xi^k(t))$ to $(SCSCLP^*)$ that is piecewise linear with partition P^k, whose solution value converges to that of $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ and P^k satisfies (3.4).*

Proof When there are no zero length intervals in P (i.e., (3.4) holds), we let

$$\xi(t) = \begin{cases} \sum_{j=i+1}^{p} \left(\hat{\xi}\left(\frac{t_j+t_{j-1}}{2}\right) + \hat{\xi}(t_j)\right), & \text{if } t = t_i,\ i = 0, 1, \cdots,\ p-1, \\ 0, & \text{if } t = T. \end{cases}$$

For $t \in (t_{i-1},\ t_i)$, we let

$$\xi(t) = \frac{t_i - t}{t_i - t_{i-1}}\xi(t_{i-1}) + \frac{t - t_{i-1}}{t_i - t_{i-1}}\left(\xi(t_i) + \hat{\xi}(t_i)\right).$$

We note that $\xi(t)$ is monotonically increasing and right continuous (albeit discontinuous). Let $\pi(t)$ and $\eta(t)$ be the piecewise linear extensions of $\hat{\pi}$ and $\hat{\eta}$ respectively. $(\pi(t),\ \eta(t),\ \xi(t))$ is a feasible solution for $(SCSCLP^*)$ by virtue of the piecewise linearity of the problem data. Now, let us check the relationship between the solution value of the newly constructed solution of $(SCSCLP^*)$ and the original solution of $AP^*(P)$. By Proposition A.2 in the Appendix, we have

$$\begin{aligned} & -\int_0^T a(t)'\, d\pi(t) \\ = \; & -a(t)'\pi(t)\ \Big|_0^T + \int_0^T \pi(t)'\, da(t) \\ = \; & \hat{\pi}(t_0+)'a(t_0) + \sum_{i=1}^{p}\left(\frac{a(t_i) - a(t_{i-1})}{t_i - t_{i-1}}\right)' \int_{t_{i-1}}^{t_i} \pi(t)\, dt \\ = \; & \hat{\pi}(t_0+)'a(t_0) + \sum_{i=1}^{p}(\hat{\pi}(t_{i-1}+) + \hat{\pi}(t_i-))'\left(a(t_i) - a\left(\frac{t_i + t_{i-1}}{2}\right)\right). \qquad (3.13) \end{aligned}$$

Since $\eta(t)$ is piecewise linear and $b(t)$ is piecewise constant with partition P, we have

$$\int_{t_{i-1}}^{t_i} b(t)'\eta(t)\, dt = \left(\frac{t_i - t_{i-1}}{2}\right)(\hat{\eta}(t_{i-1}+) + \hat{\eta}(t_i-))'b(t_i-), \quad i = 1, \cdots, p.$$

So

$$-\int_0^T b(t)'\eta(t)\, dt = -\sum_{i=1}^{p}\left(\frac{t_i - t_{i-1}}{2}\right)(\hat{\eta}(t_{i-1}+) + \hat{\eta}(t_i-))'b(t_i-) \tag{3.14}$$

Direct calculation gives

$$\begin{aligned}
& -\int_0^T h(t)'\, d\xi(t) \\
= \; & -h(t)'\xi(t)\,|_0^T + \int_0^T \xi(t)'\, dh(t) \\
= \; & h(t_0)'\xi(t_0) + \sum_{i=1}^{p}\left(\frac{h(t_i) - h(t_{i-1})}{t_i - t_{i-1}}\right)' \int_{t_{i-1}}^{t_i} \xi(t)\, dt \\
= \; & h(t_0)' \sum_{j=1}^{p}\left(\hat{\xi}\left(\frac{t_j + t_{j-1}}{2}\right) + \hat{\xi}(t_j)\right) + \\
& \sum_{i=1}^{p}\left(\frac{h(t_i) - h(t_{i-1})}{2}\right)'\left(2\sum_{j=i+1}^{p}\left(\hat{\xi}\left(\frac{t_j + t_{j-1}}{2}\right) + \hat{\xi}(t_j)\right) + \hat{\xi}\left(\frac{t_i + t_{i-1}}{2}\right) + 2\hat{\xi}(t_i)\right) \\
= \; & \sum_{i=1}^{p}\left(\hat{\xi}(t_i)'h(t_i) + \hat{\xi}\left(\frac{t_{i-1} + t_i}{2}\right)' h\left(\frac{t_{i-1} + t_i}{2}\right)\right).
\end{aligned}$$

So,

$$-\int_0^T h(t)'\, d\xi(t) = \sum_{i=1}^{p}\left(\hat{\xi}(t_i)'h(t_i) + \hat{\xi}\left(\frac{t_{i-1} + t_i}{2}\right)' h\left(\frac{t_{i-1} + t_i}{2}\right)\right). \tag{3.15}$$

Combining (3.13), (3.14) and (3.15), we see that $(\pi(t),\ \eta(t),\ \xi(t))$ has the same solution value as $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$. This proves the first half of the theorem.

Now, suppose (3.4) does not hold for P. Since P is a purified partition, by Corollary 3.1 and Lemma 3.4, the zero length intervals in P can be located only at the breakpoints in D_1^P. So for any zero length interval $[t_{i-1},\ t_i]$ that resides on $[t_l,\ t_m]$, where t_l and t_m are two consecutive breakpoints in D_1^P, either $t_{i-1} = t_l$ or $t_i = t_m$. Let $\tau \in (0,\ 1)$. We define a new solution $(\tilde{\pi}^\tau,\ \tilde{\eta}^\tau,\ \tilde{\xi}^\tau)$ in the following way.

If $t_{i-1} = t_l$, we let

$$
\begin{aligned}
\tilde{t}^{\tau}_{i-1} &= t_{i-1} \\
\tilde{t}^{\tau}_{i} &= (1-\tau)t_i + \tau t_{i+1} \\
\tilde{t}^{\tau}_{i+1} &= t_{i+1} \\
\tilde{\pi}^{\tau}(\tilde{t}^{\tau}_{i}-) &= (1-\tau)\hat{\pi}(t_i-) + \tau\hat{\pi}(t_i+) \\
\tilde{\pi}^{\tau}(\tilde{t}^{\tau}_{i}+) &= (1-\tau)\hat{\pi}(t_i+) + \tau\hat{\pi}(t_{i+1}-) \\
\tilde{\eta}^{\tau}(\tilde{t}^{\tau}_{i}-) &= (1-\tau)\hat{\eta}(t_i-) + \tau\hat{\eta}(t_i+) \\
\tilde{\eta}^{\tau}(\tilde{t}^{\tau}_{i}+) &= (1-\tau)\hat{\eta}(t_i+) + \tau\hat{\eta}(t_{i+1}-) \\
\tilde{\xi}^{\tau}\left(\frac{\tilde{t}^{\tau}_{i} + \tilde{t}^{\tau}_{i-1}}{2}\right) &= (1-\tau)\hat{\xi}\left(\frac{t_i + t_{i-1}}{2}\right) + \tau\hat{\xi}(t_i) \\
\tilde{\xi}^{\tau}(\tilde{t}^{\tau}_{i}) &= (1-\tau)\hat{\xi}(t_i) + \tau\hat{\xi}\left(\frac{t_i + t_{i+1}}{2}\right) \\
\tilde{\xi}^{\tau}\left(\frac{\tilde{t}^{\tau}_{i} + \tilde{t}^{\tau}_{i+1}}{2}\right) &= (1-\tau)\hat{\xi}\left(\frac{t_i + t_{i+1}}{2}\right).
\end{aligned}
$$

If $t_i = t_m$, we let

$$
\begin{aligned}
\tilde{t}^{\tau}_{i-2} &= t_{i-2} \\
\tilde{t}^{\tau}_{i-1} &= (1-\tau)t_{i-1} + \tau t_{i-2} \\
\tilde{t}^{\tau}_{i} &= t_i \\
\tilde{\pi}^{\tau}(\tilde{t}^{\tau}_{i-1}-) &= (1-\tau)\hat{\pi}(t_{i-1}-) + \tau\hat{\pi}(t_{i-2}+) \\
\tilde{\pi}^{\tau}(\tilde{t}^{\tau}_{i-1}+) &= (1-\tau)\hat{\pi}(t_{i-1}+) + \tau\hat{\pi}(t_{i-1}-) \\
\tilde{\eta}^{\tau}(\tilde{t}^{\tau}_{i-1}-) &= (1-\tau)\hat{\eta}(t_{i-1}-) + \tau\hat{\eta}(t_{i-2}+) \\
\tilde{\eta}^{\tau}(\tilde{t}^{\tau}_{i-1}+) &= (1-\tau)\hat{\eta}(t_{i-1}+) + \tau\hat{\eta}(t_{i-1}-) \\
\tilde{\xi}^{\tau}\left(\frac{\tilde{t}^{\tau}_{i} + \tilde{t}^{\tau}_{i-1}}{2}\right) &= (1-\tau)\hat{\xi}\left(\frac{t_i + t_{i-1}}{2}\right) + \tau\hat{\xi}(t_{i-1}) \\
\tilde{\xi}^{\tau}(\tilde{t}^{\tau}_{i-1}) &= (1-\tau)\hat{\xi}(t_{i-1}) + \tau\hat{\xi}\left(\frac{t_{i-1} + t_{i-2}}{2}\right) \\
\tilde{\xi}^{\tau}\left(\frac{\tilde{t}^{\tau}_{i-1} + \tilde{t}^{\tau}_{i-2}}{2}\right) &= (1-\tau)\hat{\xi}\left(\frac{t_{i-1} + t_{i-2}}{2}\right).
\end{aligned}
$$

For all the other quantities not defined in the above cases, we let $\tilde{t}^{\tau}_{j} = t_j$, $\tilde{\pi}^{\tau}(\tilde{t}^{\tau}_{j}-) =$

$\hat{\pi}(t_j-)$, $\tilde{\pi}^\tau(\tilde{t}_j^\tau+) = \hat{\pi}(t_j+)$, $\tilde{\eta}^\tau(\tilde{t}_j^\tau-) = \hat{\eta}(t_j-)$, $\tilde{\eta}^\tau(\tilde{t}_j^\tau+) = \hat{\eta}(t_j+)$, $\tilde{\xi}^\tau(\tilde{t}_{j-1}^\tau) = \hat{\xi}(t_{j-1})$ and $\tilde{\xi}^\tau\left(\frac{\tilde{t}_j^\tau+\tilde{t}_{j-1}^\tau}{2}\right) = \hat{\xi}\left(\frac{t_j+t_{j-1}}{2}\right)$.

Let P^τ be the partition defined from $\tilde{t}^\tau$. It is easy to check the feasibility of $(\tilde{\pi}^\tau,\ \tilde{\eta}^\tau,\ \tilde{\xi}^\tau)$ to $AP^*(P^\tau)$. Since $(\tilde{\pi}^\tau,\ \tilde{\eta}^\tau,\ \tilde{\xi}^\tau)$ converges to $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ and $\tilde{t}^\tau$ converges to $\hat{t}$ as τ tends to zero, we see that the solution value of $(\tilde{\pi}^\tau,\ \tilde{\eta}^\tau,\ \tilde{\xi}^\tau)$ in $AP^*(P^\tau)$ converges to the solution value of $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ in $AP^*(P)$. Furthermore, (3.4) holds for P^τ. Applying the first half of the theorem to P^τ, we conclude that the theorem is true for P. □

We may now summarize the relationship among the values of various discrete approximations in the following theorem (see also Theorem 3.5 in Pullan [77]).

Theorem 3.4 *For any partitions P and Q,*

$$V(AP(P)) = V(AP^*(P)) \leq V((SCSCLP^*)) \leq V((SCSCLP)) \leq V(DP(Q)).$$

Proof By strong duality result for finite dimensional linear programming, the value of the optimal solution to $AP(P)$ is the value of the optimal solution to its dual $AP^*(P)$. By Theorem 3.3, the solution value of this solution can be closely approximated by a sequence of feasible solutions to $(SCSCLP^*)$. It then follows that this value is a lower bound on $V((SCSCLP^*))$, and thus a lower bound on $V((SCSCLP))$ by Proposition 2.6. The final inequality follows from the definition of $DP(Q)$. □

Corollary 3.2 *For any partitions P and Q, if*

$$V(AP(P)) \geq V(QP(|Q|)),$$

then the optimal solution value of $QP(|Q|)$ gives the optimal solution value to $(SCSCLP)$. In particular, if a solution $(\hat{v},\ \hat{y},\ \hat{t})$ is feasible for $QP(|Q|)$ and has the same cost as the optimal value of $AP(P)$, then $(\hat{v},\ \hat{y},\ \hat{t})$ gives the optimal solution value for $(SCSCLP)$ which can be closely approximated by a sequence of feasible solutions to $(SCSCLP)$.

Proof By Lemma 3.2, the solution value of any feasible solution to $QP(|Q|)$ is an upper bound on $V((SCSCLP))$, the result follows directly from Theorem 3.4. □

3.5.2 The Doubling of Breakpoints

Based on a new discrete approximation of $(SCLP)$ similar to $AP1(P)$, Pullan [77] found a descent solution for $(SCLP)$ (consequently, a descent direction can be constructed) by patching together the current solution and a solution that has a better solution value in $AP1(P)$ than the current solution. The new solution has a strictly improved solution value in $(SCLP)$, but usually has three times as many constant control pieces as the original solution. In the following, we give a construction for a feasible solution to $(SCSCLP)$ that produces, at most, approximately twice as many breakpoints as the original feasible solution.

Let P be a partition of $[0,\ T]$, and define a new partition as follows:

$$\bar{P} = \{t_0, t_0, t_1, t_1, \cdots, t_i, t_i, t_i, \cdots, t_p, t_p\},$$

where each breakpoint in D^P has two duplicates and all the other breakpoints have only one duplicate. Intuitively, we have placed a zero length interval at the beginning of every breakpoint of P and put a zero length interval at the end of each breakpoint in D^P. Under this configuration, the set of intervals in $\bar{P}$ is the union of the intervals in P and a set of zero length intervals. We let $\bar{t}_i$ denote the $i+1$st elements in $\bar{P}$. $D_1^{\bar{P}}$ is the set of breakpoints in $\bar{P}$ that correspond to the breakpoints in D_1^P. For the i-th interval (i.e., $[t_{i-1},\ t_i]$) in P, we have a corresponding interval $[\bar{t}_{j-1},\ \bar{t}_j]$ in $\bar{P}$, where $\bar{t}_{j-1} = t_{i-1}$ and $\bar{t}_j = t_i$. We call this interval in $\bar{P}$ an old interval. For all the other intervals in $\bar{P}$, we call them new intervals. Note that all the new intervals have zero length but not vice versa.

Given a solution $(\hat{v},\ \hat{y},\ \hat{t})$ to $QP(|P|)$, we first construct a feasible solution $(\bar{v},\ \bar{y},\ \bar{t})$ to $QP(|\bar{P}|)$ and then show a descent direction for this solution in $QP(|\bar{P}|)$, although we need not use the same direction in the new algorithm as the one constructed in the proof. This solution has the same solution value in $QP(|\bar{P}|)$ as the current solution in $QP(|P|)$ and has approximately twice as many intervals, fewer than the one constructed by Pullan [77].

Let $(\hat{v},\ \hat{y},\ \hat{t})$ be a feasible solution for $QP(|P|)$. For the i-th interval in $\bar{P}$, if it is an old interval, we let interval j be the corresponding interval in P and set

$$\bar{v}(\bar{t}_i) = \hat{v}(t_j), \qquad \bar{y}(\bar{t}_i) = \hat{y}(t_j). \tag{3.16}$$

We let $\bar{v}(\bar{t}_i) = 0$ if interval i in $\bar{P}$ is a new interval and let $\bar{y}(\bar{t}_i) = \hat{y}(t_j)$, where j is the interval in P that corresponds to the closest old interval in $\bar{P}$ to the left of $[t_{i-1},\ t_i]$ (with the convention that $\bar{y}(\bar{t}_1) = y(t_0)$ and $\bar{y}(\bar{t}_0) = y(t_0)$).

It is easy to verify that $(\bar{v},\ \bar{y},\ \bar{t})$ is feasible to $QP(|\bar{P}|)$ and has the same solution value in $QP(|\bar{P}|)$ as $(\hat{v},\ \hat{y},\ \hat{t})$ in $QP(|P|)$.

3.5.3 A Descent Direction

Let $\Phi(v, y, t)$ be the solution value in $AP(P)$ for a feasible solution $(v,\ y,\ t)$ to $AP(P)$ and let $\Psi(v, y, t)$ be the solution value in $QP(|P|)$ for a feasible solution $(v,\ y,\ t)$ for $QP(|P|)$. According to Corollary 3.2, a feasible solution $(\hat{v},\ \hat{y},\ \hat{t})$ to $QP(|P|)$ gives the optimal solution value of $(SCSCLP)$ if the optimal solution to $AP(P)$ has an equal or larger solution value. If so, we can stop the algorithm. Otherwise, there exists $(\bar{\bar{v}},\ \bar{\bar{y}},\ \bar{\bar{t}})$ feasible for $AP(P)$ and has a strictly smaller solution value in $AP(P)$, i.e., we have

$$\delta \stackrel{\text{def}}{=} \Phi(\bar{\bar{v}}, \bar{\bar{y}}, \bar{\bar{t}}) - \Psi(\hat{v},\ \hat{y},\ \hat{t}) < 0. \tag{3.17}$$

Note that $|\delta|$ is an upper bound on the duality gap between $(SCSCLP)$ and $(SCSCLP^*)$.

Let $\epsilon \in [0,\ 1]$. For every interval $[t_{i-1},\ t_i]$, we define

$$\epsilon_i = \frac{(t_i - t_{i-1})\epsilon}{2}.$$

We define a new partition P^ϵ of $[0,\ T]$ as follows.

$$P^\epsilon \stackrel{\text{def}}{=} \{t_0, t_0 + \epsilon_1, t_1 - \epsilon_1, t_1 + \epsilon_2, \cdots, t_i - \epsilon_i, t_i, t_i + \epsilon_{i+1}, \cdots, t_p - \epsilon_p, t_p\},$$

where the breakpoint t_i in $P \setminus D_1^P$ is replaced by two elements $t_i - \epsilon_i$ and $t_i + \epsilon_{i+1}$, for breakpoint t_i in D^P, we add two elements $t_i - \epsilon_i$ and $t_i + \epsilon_{i+1}$ and we add $t_0 + \epsilon_1$ and $t_p - \epsilon_p$ for t_0 and t_p respectively. We define the vector t^ϵ from P^ϵ by mapping t_i^ϵ to the $i + 1$st element in P^ϵ. We construct a descent solution $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ with partition P^ϵ as follows.

When P does not have any zero length intervals, let $\bar{u}(t)$, $\hat{u}(t)$ and $\hat{\bar{u}}(t)$ be the piecewise constant extension of $\tilde{\bar{u}}$, $\hat{u}$ and $\hat{\bar{u}}$ respectively, where $\tilde{\bar{u}}$ is defined from $\tilde{\bar{v}}$ by

$$\tilde{\bar{u}}(t_{i-1}+) = 2\frac{\tilde{\bar{v}}(t_{i-1}+)}{t_i - t_{i-1}}, \quad \tilde{\bar{u}}(t_i-) = 2\frac{\tilde{\bar{v}}(t_i-)}{t_i - t_{i-1}},$$

$\hat{u}$ is defined from $\hat{v}$ by (3.3) and $\hat{\bar{u}}$ is defined as

$$\hat{\bar{u}}(t_{i-1}+) = \frac{\tilde{\bar{v}}(t_{i-1}+) + \tilde{\bar{v}}(t_i-)}{t_i - t_{i-1}}.$$

We construct the new control by patching together $\tilde{\bar{u}}(t)$, $\hat{u}(t)$ and $\hat{\bar{u}}(t)$ as follows

$$u^\epsilon(t) = \begin{cases} \tilde{\bar{u}}(t), & t \in [t_{i-1},\ t_{i-1} + \epsilon_i) \cup [t_i - \epsilon_i,\ t_i),\ t_i \in D^P \\ \tilde{\bar{u}}(t), & t \in [t_{p-1},\ t_{p-1} + \epsilon_p) \cup [t_p - \epsilon_p,\ t_p], \\ \hat{u}(t), & t \in [t_{i-1} + \epsilon_i,\ t_i - \epsilon_i), \\ \hat{\bar{u}}(t), & \text{otherwise.} \end{cases} \tag{3.18}$$

Having constructed the control, the construction of the states for $(SCLP)$ is straight forward. We will see how this generalize to $(SCSCLP)$ shortly. Our construction of a descent solution $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ for $(\hat{v},\ \hat{y},\ \hat{t})$ is illustrated in Figure 3-2.

However, if $t_{i-1} = t_i$ for some i, the u variables in the previous paragraph are not properly defined. Fortunately, we can bypass this difficulty by working on v variables. We define v^ϵ as follows. Let t_l and t_m be two consecutive breakpoints in D_1^P. Let $[t_i + \epsilon_{i+1},\ t_{i+1} - \epsilon_{i+1}]$ and $[t_{i+1} - \epsilon_{i+1},\ t_{i+1} + \epsilon_{i+2}]$ be two intervals that reside on $[t_l,\ t_m]$. If t_j^ϵ is the breakpoint in P^ϵ that is mapped to $t_{i+1} - \epsilon_{i+1}$, we let

$$\begin{aligned} v^\epsilon(t_j^\epsilon) &= (1-\epsilon)\hat{v}(t_{i+1}) \\ v^\epsilon(t_{j+1}^\epsilon) &= \epsilon(\tilde{\bar{v}}(t_{i+1}-) + \tilde{\bar{v}}(t_{i+1}+)). \end{aligned} \tag{3.19}$$

If t_j^ϵ is the breakpoint in P^ϵ that is mapped to t_l, we let

$$v^\epsilon(t_{j+1}^\epsilon) = \epsilon\tilde{\bar{v}}(t_l+). \tag{3.20}$$

If t_j^ϵ is the breakpoint in P^ϵ that is mapped to t_m, we let

$$v^\epsilon(t_j^\epsilon) = \epsilon\tilde{\bar{v}}(t_m-). \tag{3.21}$$

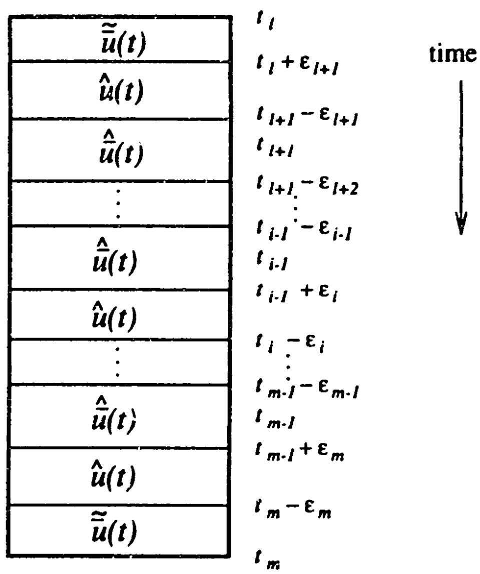

Figure 3-2: The construction of a descent solution
where t_l and t_m are two consecutive breakpoints in D_1.

We define y^ϵ in three different cases as follows. For the breakpoint t_j^ϵ in P^ϵ that is mapped to $t_{i-1} + \epsilon_i$, we let

$$y^\epsilon(t_j^\epsilon) = (1-\epsilon)\hat{y}(t_{i-1}) + \epsilon\tilde{\tilde{y}}\left(\frac{t_i + t_{i-1}}{2}\right). \tag{3.22}$$

For the breakpoint t_j^ϵ in P^ϵ that is mapped to $t_i - \epsilon_i$, we let

$$y^\epsilon(t_j^\epsilon) = (1-\epsilon)\hat{y}(t_i) + \epsilon\tilde{\tilde{y}}\left(\frac{t_i + t_{i-1}}{2}\right). \tag{3.23}$$

For the breakpoint t_j^ϵ in P^ϵ that is mapped to t_i, we let

$$y^\epsilon(t_j^\epsilon) = (1-\epsilon)\hat{y}(t_i) + \epsilon\tilde{\tilde{y}}(t_i). \tag{3.24}$$

When ϵ is small, $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ is a descent solution as shown in the following theorem.

Theorem 3.5 *If (3.17) holds, then $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ is a feasible solution to $QP(|\bar{P}|)$ and*

$$\Psi(v^\epsilon,\ y^\epsilon,\ t^\epsilon) - \Psi(\bar{v},\ \bar{y},\ \bar{t}) = \epsilon\delta - o(\epsilon),$$

where δ is defined in (3.17). *For ϵ small enough, $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ has a strictly smaller solution value than $(\bar{v},\ \bar{y},\ \bar{t})$.*

Proof The feasibility of $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ easily follows.

By definition, we have

$$\Psi(\bar{v},\ \bar{y},\ \bar{t}) = \Psi(\hat{v},\ \hat{y},\ \hat{t}) = \sum_{i=1}^{p} \hat{v}(t_i)'c(\frac{t_i + t_{i-1}}{2}) + \sum_{i=1}^{p} \frac{t_i - t_{i-1}}{2}(\hat{y}(t_i) + \hat{y}(t_{i-1}))'g(t_{i-1}+),$$

$$\Psi(v^\epsilon,\ y^\epsilon,\ t^\epsilon) = \sum_{i=1}^{|P^\epsilon|-1} c(\frac{t_i^\epsilon + t_{i-1}^\epsilon}{2})'v^\epsilon(t_i^\epsilon) + \sum_{i=1}^{|P^\epsilon|-1} \frac{t_i^\epsilon - t_{i-1}^\epsilon}{2}(y^\epsilon(t_i^\epsilon) + y^\epsilon(t_{i-1}^\epsilon))'g(t_{i-1}^\epsilon+).$$

Let t_l and t_m be two consecutive breakpoints in D_1^P and let $t_{\bar{l}}^\epsilon$ and $t_{\bar{m}}^\epsilon$ be the corresponding breakpoints in $D_1^{P^\epsilon}$. We have

$$\begin{aligned}
& \sum_{i=l+1}^{m} \hat{v}(t_i)'c(\frac{t_i + t_{i-1}}{2}) - \sum_{i=\bar{l}+1}^{\bar{m}} c(\frac{t_i^\epsilon + t_{i-1}^\epsilon}{2})'v^\epsilon(t_i^\epsilon) \\
= \ & \sum_{i=l+1}^{m} \hat{v}(t_i)'c(\frac{t_i + t_{i-1}}{2}) - \sum_{i=l+1}^{m} (1-\epsilon)\hat{v}(t_i)'c(\frac{t_i + t_{i-1}}{2}) - \\
& \sum_{i=l+1}^{m-1} \epsilon(\tilde{\bar{v}}(t_i+) + \tilde{\bar{v}}(t_i-))'c(t_i + \frac{\epsilon_{i+1} - \epsilon_i}{2}) - \\
& \epsilon\tilde{\bar{v}}(t_l+)'c(t_l + \frac{\epsilon_{l+1}}{2}) - \epsilon\tilde{\bar{v}}(t_m-)'c(t_m - \frac{\epsilon_m}{2}) \\
= \ & \sum_{i=l+1}^{m} \epsilon(\hat{v}(t_i)'c(\frac{t_i + t_{i-1}}{2}) - (\tilde{\bar{v}}(t_{i-1}+)'c(t_{i-1}+) + \tilde{\bar{v}}(t_i-)'c(t_i-))) + o(\epsilon). \quad (3.25)
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{i=l+1}^{m} \frac{t_i - t_{i-1}}{2}(\hat{y}(t_i) + \hat{y}(t_{i-1}))'g(t_{i-1}+) - \sum_{i=\bar{l}+1}^{\bar{m}} \frac{t_i^\epsilon - t_{i-1}^\epsilon}{2}(y^\epsilon(t_i^\epsilon) + y^\epsilon(t_{i-1}^\epsilon))'g(t_{i-1}^\epsilon+) \\
= \ & \sum_{i=l+1}^{m} \frac{t_i - t_{i-1}}{2}(\hat{y}(t_i) + \hat{y}(t_{i-1}))'g(t_{i-1}+) -
\end{aligned}$$

$$
\begin{aligned}
& \sum_{i=l+1}^{m} (1-\epsilon)\frac{t_i - t_{i-1}}{2}\left((1-\epsilon)(\hat{y}(t_i)+\hat{y}(t_{i-1})) + 2\epsilon\bar{\bar{y}}\left(\frac{t_i+t_{i-1}}{2}\right)\right)' g(t_{i-1}+) - \\
& \sum_{i=l+1}^{m} \frac{\epsilon_i + \epsilon_{i+1}}{2}\left(2(1-\epsilon)\hat{y}(t_i) + \epsilon\left(\bar{\bar{y}}\left(\frac{t_i+t_{i-1}}{2}\right) + \bar{\bar{y}}\left(\frac{t_{i+1}+t_i}{2}\right)\right)\right)' g(t_{i-1}+) - \\
& \frac{\epsilon_{l+1}}{2}\left(2(1-\epsilon)\hat{y}(t_l) + \epsilon\left(\bar{\bar{y}}(t_l) + \bar{\bar{y}}\left(\frac{t_{l+1}+t_l}{2}\right)\right)\right)' g(t_l+) - \\
& \frac{\epsilon_m}{2}\left(2(1-\epsilon)\hat{y}(t_m) + \epsilon\left(\bar{\bar{y}}(t_m) + \bar{\bar{y}}\left(\frac{t_{m-1}+t_m}{2}\right)\right)\right)' g(t_m-) \\
= & \sum_{i=l+1}^{m} \epsilon\frac{t_i - t_{i-1}}{2}(\hat{y}(t_i)+\hat{y}(t_{i-1}))' g(t_{i-1}+) - \\
& \sum_{i=l+1}^{m} (t_i - t_{i-1})\epsilon\bar{\bar{y}}\left(\frac{t_i+t_{i-1}}{2}\right)' g(t_{i-1}+) + o(\epsilon). \qquad (3.26)
\end{aligned}
$$

Summing up (3.25) and (3.26) over all pairs of consecutive breakpoints in D_1, we have

$$
\begin{aligned}
& \Psi(v^\epsilon,\ y^\epsilon,\ t^\epsilon) - \Psi(\bar{v},\ \bar{y},\ \bar{t}) \\
= & \Psi(v^\epsilon,\ y^\epsilon,\ t^\epsilon) - \Psi(\hat{v},\ \hat{y},\ \hat{t}) \\
= & \sum_{i=1}^{p} \epsilon(\bar{\bar{v}}(t_{i-1}+)'c(t_{i-1}+) + \bar{\bar{v}}(t_i-)'c(t_i-) - \hat{v}(t_i)'c(\frac{t_i+t_{i-1}}{2})) + \\
& \epsilon\sum_{i=1}^{p}\left((t_i - t_{i-1})\bar{\bar{y}}\left(\frac{t_i+t_{i-1}}{2}\right)' g(t_{i-1}+) - \frac{t_i - t_{i-1}}{2}(\hat{y}(t_i)+\hat{y}(t_{i-1}))' g(t_{i-1}+)\right) + o(\epsilon) \\
= & \epsilon(\Phi(\bar{\bar{v}},\bar{\bar{y}},\bar{\bar{t}}) - \Psi(\hat{v},\ \hat{y},\ \hat{t})) + o(\epsilon) \\
= & \epsilon\delta + o(\epsilon) \qquad (3.27)
\end{aligned}
$$

So, when ϵ is small enough, $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ is a strictly improved feasible solution to $QP(|\bar{P}|)$.

□

Interestingly, the new solution $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ gives a descent direction for $(\bar{v},\ \bar{y},\ \bar{t})$ in $QP(|\bar{P}|)$. This solution can also be used to show that the first Frank-Wolfe iterate for $(\bar{v},\ \bar{y},\ \bar{t})$ provides an upper bound on the current duality gap, as we now illustrate.

Let $[t_l,\ t_m]$ be two consecutive breakpoints in D_1^P. We define a new partition $\bar{\bar{P}}$ as follows. The set of breakpoints of $\bar{\bar{P}}$ that resides on $[t_l,\ t_m]$ is $\{t_l,\ \frac{t_l+t_{l+1}}{2},\ \frac{t_l+t_{l+1}}{2},\cdots,\ t_m\}$, i.e., the union of $\{t_l,\ t_m\}$ with the set of midpoints of the intervals in P, and each midpoint appears exactly twice. We construct $(\bar{\bar{v}},\ \bar{\bar{y}},\ \bar{\bar{t}})$ as follows. The set of breakpoints of $(\bar{\bar{v}},\ \bar{\bar{y}},\ \bar{\bar{t}})$ is $\bar{\bar{P}}$. Let

$$\bar{\bar{v}}_j = \begin{cases} \bar{v}(t_{i+1}-) + \bar{v}(t_{i+1}+); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [\frac{t_i+t_{i+1}}{2},\ \frac{t_{i+1}+t_{i+2}}{2}] \\ \bar{v}(t_l+); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [t_l,\ \frac{t_l+t_{l+1}}{2}] \\ \bar{v}(t_m-); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [\frac{t_{m-1}+t_m}{2},\ t_m] \\ 0; & \text{otherwise,} \end{cases}$$

$\bar{\bar{y}}(\bar{\bar{t}}_0) = y(t_0)$, and

$$\bar{\bar{y}}(\bar{\bar{t}}_j) = \begin{cases} \bar{y}\left(\frac{t_{i+1}+t_{i+2}}{2}\right); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [\frac{t_i+t_{i+1}}{2},\ \frac{t_{i+1}+t_{i+2}}{2}] \\ \bar{y}\left(\frac{t_l+t_{l+1}}{2}\right); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [t_l,\ \frac{t_l+t_{l+1}}{2}] \\ \bar{y}(t_m); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [\frac{t_{m-1}+t_m}{2},\ t_m] \\ \bar{y}\left(\frac{t_{i+1}+t_i}{2}\right); & \text{if the } j\text{-th interval of } \bar{\bar{P}} \text{ is } [\frac{t_i+t_{i+1}}{2},\ \frac{t_i+t_{i+1}}{2}]. \end{cases}$$

Theorem 3.6 *For $\epsilon \in [0,\ 1]$, let t^ϵ be defined by P^ϵ. Let $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ be the solution to $QP(|\bar{P}|)$ defined by (3.19)-(3.24), we have*

$$\begin{aligned} v^\epsilon &= \epsilon\bar{\bar{v}} + (1-\epsilon)\,\bar{v} \\ y^\epsilon &= \epsilon\bar{\bar{y}} + (1-\epsilon)\,\bar{y} \\ t^\epsilon &= \epsilon\bar{\bar{t}} + (1-\epsilon)\,\bar{t} \end{aligned}$$

and $(\bar{\bar{v}},\ \bar{\bar{y}},\ \bar{\bar{t}})$ is feasible for $QP(|\bar{P}|)$.

Proof This is the direct consequence of the definition of $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$ and $(\bar{\bar{v}},\ \bar{\bar{y}},\ \bar{\bar{t}})$. □

If we pick $(\bar{\bar{v}}, \bar{\bar{y}},\ \bar{\bar{t}})$, introduced in (3.17), an optimal solution for $AP(P)$, by Theorem 3.4, $|\delta|$ is an upper bound on the current duality gap. By (3.27) and Theorem 3.6, the negative objective value of the first Frank-Wolfe iterate (cf. $(SUBLP^k)$) for $(\bar{v},\ \bar{y},\ \bar{t})$ gives an upper bound on the current duality gap.

3.6 A New Algorithm for $(SCSCLP)$

In this section, we give a generic Successive Quadratic Programming algorithm for problem $(SCSCLP)$.

Algorithm $\mathcal{A}$ $(E, F, G, H, a(t), b(t), c(t), g(t), h(t), T, \beta)$.

$k = 0$. Let d be the current duality gap initially set to infinity.

Let $(v^k,\ y^k,\ t^k)$ be a feasible solution to $QP(|P^0|)$. Let P^0 be a partition on $[0,\ T]$ such that $a(t)$, $c(t)$ and $h(t)$ are piecewise linear with P^0 and $b(t)$ and $g(t)$ are piecewise constant with P^0.

while $d > \beta$ do

1. Calculate a KKT point of $QP(|P^k|)$ which has an equal or better solution value than $(v^k,\ y^k,\ t^k)$.

2. Recursively remove redundant intervals in P^k as follows.
   Apply Procedure $PURIFY$ to all pairs of consecutive breakpoints in $D_1^{P^k}$. Let $(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$ be the resulting solution and let Q be the resulting partition. If $(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$ is not a KKT point of $QP(|Q|)$, let $(v^k,\ y^k,\ t^k) = (\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$ and $P^k = Q$ and go to step 1. Otherwise, we denote the resulting purified partition as $\tilde{P}^k = \{t_0, t_1, \cdots, t_p\}$.

3. Double the number of intervals. Define P^{k+1} as

$$P^{k+1} = \{t_0, t_0, t_1, t_1, \cdots, t_i, t_i, t_i, \cdots, t_p, t_p\},$$

   where each breakpoint in D has two duplicates and all the other breakpoints have only one duplicate. Construct a feasible solution $(\bar{v}^{k+1},\ \bar{y}^{k+1},\ \bar{t}^{k+1})$ for $QP(|P^{k+1}|)$ as in (3.16).

4. Calculate the current duality gap d. If the solution value of $(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$ is the same as the optimal value of $AP(\tilde{P}^k)$, stop the algorithm. Otherwise

5. Get a strictly improved solution $(v^{k+1},\ y^{k+1},\ t^{k+1})$ from $(\bar{v}^{k+1},\ \bar{y}^{k+1},\ \bar{t}^{k+1})$ for $QP(|P^{k+1}|)$.

6. Let $k = k + 1$.

end while

Remarks:

1. In step 1 of Algorithm $\mathcal{A}$, we can use the Frank-Wolfe method or general matrix splitting algorithms to compute a KKT point of $QP(|P^k|)$.

2. Algorithm $\mathcal{A}$ will not loop between step 1 and step 2 forever, because every time Algorithm $\mathcal{A}$ goes from step 2 to step 1, the cardinality of P^k is reduced at least by 1.

3. In step 4 of Algorithm $\mathcal{A}$, we can let $d = V(QP(|\tilde{P}^k|)) - V(AP(\tilde{P}^k))$. We can also let d be the negative objective value of the first Frank-Wolfe iterate for $(\bar{v},\ \bar{y},\ \bar{t})$ and so, instead of checking whether the solution value of $(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$ is the same as the optimal value of $AP(\tilde{P}^k)$, we can check whether the objective value of the first Frank-Wolfe iterate for $(\bar{v},\ \bar{y},\ \bar{t})$ is zero.

4. In step 5 of Algorithm $\mathcal{A}$, we can use the direction constructed in Section 3.5.3 (cf. $(v^\epsilon,\ y^\epsilon,\ t^\epsilon)$). We can also use the Frank-Wolfe method or general matrix splitting algorithms to find a descent direction for $(\bar{v}^{k+1},\ \bar{y}^{k+1},\ \bar{t}^{k+1})$. By Theorem 3.5, we are guaranteed find a descent direction.

3.7 Convergence of the New Algorithm

The theory behind the new algorithm is the following. We can use the Frank-Wolfe method or general matrix splitting algorithms to compute a series of KKT points to a series of generally nonconvex quadratic programs. These KKT points have nondecreasing solution values. By Corollary 3.2, we can detect whether a KKT point gives an optimal solution to $(SCSCLP)$. If it does, we terminate the algorithm. If not, by Theorem 3.5, we can find a new solution with approximately twice as many constant control pieces as the current solution but with a strictly improved cost. Since there are only a finite number of different solution values for the KKT points of every quadratic program constructed, and there is an upper bound on the size of the quadratic programs we encounter (see more elaboration later), a finite convergence result readily follows. Based on the primal solution, we can compute an optimal dual solution for $(SCSCLP^*)$.

Contrary to the convergence analysis of a variety of algorithms for $(SCLP)$, we do not need to let the norm of the maximal length interval in the discretization tend to zero (as in Pullan [77]), we do not need the explicit knowledge of all the extreme points of certain set of finite dimensional linear programs either (as in Anderson and Nash [2]). Most importantly, we prove the absence of a duality gap result as a byproduct of the new algorithm, even when there is no optimal solution for $(SCSCLP)$.

In the following, we give upper bounds on the cardinality of $\tilde{P}^k$, the purified partition in step 2 of Algorithm $\mathcal{A}$. Since by Lemma 3.4 and Corollary 3.1, we know the total number of zero length intervals in $\tilde{P}^k$ is at most $2|D_1^{\tilde{P}^k}|$, we only need to bound the number of positive length intervals in $\tilde{P}^k$. We map each positive length interval of $\tilde{P}^k$ to an extreme point of certain set of linear programs and then show that the mapping is injective. Before doing so, we give some more notations and several useful lemmas.

Let t_l and t_m be two consecutive breakpoints in $D_1^{\tilde{P}^k}$. By definition, $a(t)$, $c(t)$ and $h(t)$ are linear and $b(t)$ and $g(t)$ are constant on $[t_l,\ t_m)$. Let $[\tilde{t}_{i-1},\ \tilde{t}_i]$ and $[\tilde{t}_i,\ \tilde{t}_{i+1}]$ be two adjacent positive length intervals in partition $\tilde{P}^k$, such that $[\tilde{t}_{i-1},\ \tilde{t}_{i+1}] \subseteq [t_l,\ t_m]$. Let $\Delta t_i = \tilde{t}_i - \tilde{t}_{i-1}$ and $\Delta t_{i+1} = \tilde{t}_{i+1} - \tilde{t}_i$. We have $\Delta t_i > 0$ and $\Delta t_{i+1} > 0$ by assumption. Let $(\tilde{v}^k, \tilde{y}^k, \tilde{t}^k)$ be the resulting solution in step 2 of Algorithm $\mathcal{A}$. Let J_i be the set of indices of the constraints in $F\tilde{y}^k(\tilde{t}_i) \leq h(\tilde{t}_i)$ that are binding. Let

$$\tilde{u}^k(\tilde{t}_{i-1}+) = \frac{\tilde{v}^k(\tilde{t}_i)}{\Delta t_i}$$

and

$$\tilde{u}^k(\tilde{t}_i+) = \frac{\tilde{v}^k(\tilde{t}_{i+1})}{\Delta t_{i+1}}.$$

It is obvious that $(\tilde{u}^k(\tilde{t}_i+),\ \frac{\tilde{y}^k(\tilde{t}_{i+1})-\tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}})$ is a feasible solution to the following linear system:

$$(SYS_{J_i}) \qquad \begin{aligned} & G\tilde{u}^k(\tilde{t}_i+) + E\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} = \dot{a}(\tilde{t}_i) \\ & H\tilde{u}^k(\tilde{t}_i+) \leq b(\tilde{t}_i) \\ & \left(F\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}}\right)_{J_i} \leq \dot{h}(\tilde{t}_i) \end{aligned}$$

$$\tilde{u}^k(\tilde{t}_i+) \geq 0.$$

By introducing new variables, we can eliminate $\frac{\tilde{y}^k(\tilde{t}_{i+1})-\tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}}$ in (SYS_{J_i}) and transform (SYS_{J_i}) into a linear program in standard form as follows:

$$(SYS1_{J_i}) \qquad \begin{aligned} & G\tilde{u}^k(\tilde{t}_i+) + E(w_{i+1} - w_i) = \dot{a}(\tilde{t}_i) \\ & H\tilde{u}^k(\tilde{t}_i+) + \tilde{z}^k(\tilde{t}_i+) = b(\tilde{t}_i) \\ & (F(w_{i+1} - w_i))_{J_i} + x = \dot{h}(\tilde{t}_i) \\ & x \geq 0, w_{i+1} \geq 0, w_i \geq 0, \tilde{u}^k(\tilde{t}_i+) \geq 0, \tilde{z}^k(\tilde{t}_i+) \geq 0. \end{aligned}$$

Every extreme point of the new linear program $(SYS1_{J_i})$ defines a unique feasible solution to (SYS_{J_i}), which is called a generalized extreme point for (SYS_{J_i}). Every extreme ray of the new linear program $(SYS1_{J_i})$ defines a unique ray to (SYS_{J_i}), which is called a generalized extreme ray for (SYS_{J_i}). Since $(SYS1_{J_i})$ is a feasible finite dimensional linear program in standard form, the resolution theorem applies. After translating the result into variables in (SYS_{J_i}), we have the following analog of the resolution theorem for (SYS_{J_i}).

Lemma 3.5 *Every feasible solution of (SYS_{J_i}) can be written as the sum of a convex combination of the generalized extreme points of (SYS_{J_i}) and a linear combination (with nonnegative coefficients) of generalized extreme rays to (SYS_{J_i}).*

By Lemma 3.5, we have

$$\begin{aligned} \tilde{u}^k(\tilde{t}_i+) &= \sum_{j=1}^{k^{(i)}} \lambda_j^{(i)} s_j^{(i)} + \sum_{j=1}^{q^{(i)}} \mu_j^{(i)} r_j^{(i)}, \\ \frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} &= \sum_{j=1}^{k^{(i)}} \lambda_j^{(i)} \bar{s}_j^{(i)} + \sum_{j=1}^{q^{(i)}} \mu_j^{(i)} \bar{r}_j^{(i)} \end{aligned} \qquad (3.28)$$

for some positive $k^{(i)} \geq 1$, where $\lambda_j^{(i)} > 0$, $\sum_{j=1}^{k^{(i)}} \lambda_j^{(i)} = 1$, $\mu_j^{(i)} > 0$, $(s_j^{(i)}, \bar{s}_j^{(i)})$s are generalized extreme points to system (SYS_{J_i}) and $(r_j^{(i)}, \bar{r}_j^{(i)})$s are generalized extreme rays to system (SYS_{J_i}). WLOG, assume that we have sorted $(s_j^{(i)}, \bar{s}_j^{(i)})$ in the following order

$$\dot{c}(\tilde{t}_i)' s_j^{(i)} - g(t_l+)' \bar{s}_j^{(i)} \geq \dot{c}(\tilde{t}_i)' s_{j+1}^{(i)} - g(t_l+)' \bar{s}_{j+1}^{(i)} \qquad \text{for all } j. \tag{3.29}$$

We have the following result on $(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$.

Lemma 3.6

$$\Psi(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k) \leq \Psi(v^k,\ y^k,\ t^k).$$

Proof Since Procedure $\mathcal{A}$ does not increase the solution value of the current solution, the result immediately follows. □

Lemma 3.7 *Suppose* (3.28) *and* (3.29) *hold for* $\tilde{u}^k(\tilde{t}_{i-1}+)$ *and* $\tilde{u}^k(\tilde{t}_i+)$. *Furthermore, suppose* $[\tilde{t}_{i-1},\ \tilde{t}_i]$ *is not the first positive length interval that resides on* $[t_l,\ t_m]$, *and we have*

$$\dot{c}(\tilde{t}_i)' s_1^{(i-1)} - g(t_l+)' \bar{s}_1^{(i-1)} > \dot{c}(\tilde{t}_i)' s_1^{(i)} - g(t_l+)' \bar{s}_1^{(i)}$$

for the two adjacent positive length intervals $[\tilde{t}_{i-1},\ \tilde{t}_i]$ *and* $[\tilde{t}_i,\ \tilde{t}_{i+1}]$ *that reside on* $[t_l,\ t_m]$.

Proof We first show that

$$\dot{c}(\tilde{t}_i)' r_j^{(i)} - g(t_l+)' \bar{r}_j^{(i)} \leq 0 \tag{3.30}$$

for every $j \leq q^{(i)}$ without assuming that $[\tilde{t}_{i-1},\ \tilde{t}_i]$ is not the first positive length interval that resides on $[t_l,\ t_m]$.

Let $\tau \in (0,\ 1)$. Suppose

$$\tilde{u}^k(\tilde{t}_i+) = \tau u_1 + (1-\tau) u_2$$

and

$$\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} = \tau y_1 + (1-\tau) y_2,$$

where $(u_1,\ y_1)$ and $(u_2,\ y_2)$ are two feasible solutions for (SYS_{J_i}). Let γ be the largest scalar in $(0,\ \tau\Delta t_{i+1}]$ such that $F(\tilde{y}^k(\tilde{t}_i) + \gamma y_1) \leq h(\tilde{t}_i)$. Such γ exists by virtue of the feasibility of $(u_1,\ y_1)$ and $(u_2,\ y_2)$ to system (SYS_{J_i}). For any $\Delta t \in (0,\ \gamma)$, we consider the following perturbation of $(\tilde{v}^k, \tilde{y}^k, \tilde{t}^k)$, as shown in Figure 3-3.

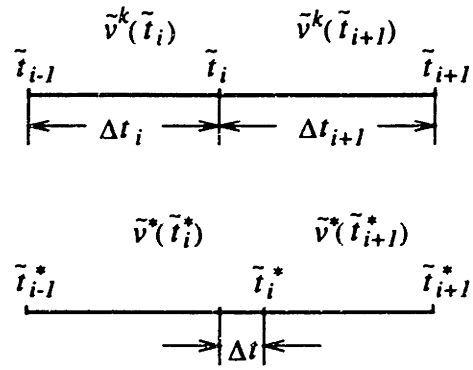

Figure 3-3: Perturbation of the solution

$$\begin{aligned}
\tilde{t}_j^* &= \begin{cases} \tilde{t}_i + \Delta t, & \text{if } j = i, \\ \tilde{t}_j, & \text{otherwise} \end{cases} \\
\tilde{v}^*(\tilde{t}_j^*) &= \begin{cases} \tilde{v}^k(\tilde{t}_i) + u_1^{(i)}\Delta t, & \text{if } j = i, \\ \tilde{v}^k(\tilde{t}_j) - u_1^{(i)}\Delta t, & \text{if } j = i+1, \\ \tilde{v}^k(\tilde{t}_j), & \text{otherwise} \end{cases} \\
\tilde{y}^*(\tilde{t}_j^*) &= \begin{cases} \tilde{y}^k(\tilde{t}_i) + y_1^{(i)}\Delta t, & \text{if } j = i, \\ \tilde{y}^k(\tilde{t}_j), & \text{otherwise} \end{cases}
\end{aligned}$$

We can easily check the feasibility of $(\tilde{v}^*,\ \tilde{y}^*,\ \tilde{t}^*)$ to $QP(|\tilde{P}|)$. So

$$\begin{aligned}
&\sum_{j=1}^{p} \tilde{v}^*(\tilde{t}_j^*)'c\left(\frac{\tilde{t}_{i-1}^* + \tilde{t}_i^*}{2}\right) - \sum_{j=1}^{p} \tilde{v}^k(\tilde{t}_j)'c\left(\frac{\tilde{t}_{i-1} + \tilde{t}_i}{2}\right) \\
= &\left(\frac{c(\tilde{t}_{i-1}) + c(\tilde{t}_i^*)}{2}\right)' \tilde{v}^*(\tilde{t}_i^*) + \left(\frac{c(\tilde{t}_i^*) + c(\tilde{t}_{i+1})}{2}\right)' \tilde{v}^*(\tilde{t}_{i+1}^*) - \\
&\left(\frac{c(\tilde{t}_{i-1}) + c(\tilde{t}_i)}{2}\right)' \tilde{v}^k(\tilde{t}_i) - \left(\frac{c(\tilde{t}_i) + c(\tilde{t}_{i+1})}{2}\right)' \tilde{v}^k(\tilde{t}_{i+1})
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{c(\tilde{t}_i^*) - c(\tilde{t}_i)}{2}\right)' \tilde{v}^k(\tilde{t}_i) + \left(\frac{c(\tilde{t}_i^*) - c(\tilde{t}_i)}{2}\right)' \tilde{v}^k(\tilde{t}_{i+1}) - \\
&\qquad \frac{\Delta t}{2}(\Delta t_i + \Delta t_{i+1})\dot{c}(\tilde{t}_i)' u_1^{(i)} \\
&= \frac{\Delta t \Delta t_i}{2}(\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i-) - \dot{c}(\tilde{t}_i)'u_1^{(i)}) + \\
&\quad \frac{\Delta t \Delta t_{i+1}}{2}(\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i+) - \dot{c}(\tilde{t}_i)'u_1^{(i)}) \qquad (3.31)
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\sum_{j=1}^{p} \frac{\tilde{t}_i^* - \tilde{t}_{i-1}^*}{2}(\tilde{y}^*(\tilde{t}_i^*) + \tilde{y}^*(\tilde{t}_{i-1}^*))'g(\tilde{t}_{i-1}^*+) - \\
&\sum_{j=1}^{p} \frac{\tilde{t}_i - \tilde{t}_{i-1}}{2}(\tilde{y}^k(\tilde{t}_i) + \tilde{y}^k(\tilde{t}_{i-1}))'g(\tilde{t}_{i-1}+) \\
&= \frac{\tilde{t}_i^* - \tilde{t}_{i-1}^*}{2}(\tilde{y}^*(\tilde{t}_i^*) + \tilde{y}^*(\tilde{t}_{i-1}^*))'g(\tilde{t}_{i-1}^*+) + \frac{\tilde{t}_{i+1}^* - \tilde{t}_i^*}{2}(\tilde{y}^*(\tilde{t}_{i+1}^*) + \tilde{y}^*(\tilde{t}_i^*))'g(\tilde{t}_i^*+) - \\
&\quad \frac{\tilde{t}_i - \tilde{t}_{i-1}}{2}(\tilde{y}^k(\tilde{t}_i) + \tilde{y}^k(\tilde{t}_{i-1}))'g(\tilde{t}_{i-1}+) - \frac{\tilde{t}_{i+1} - \tilde{t}_i}{2}(\tilde{y}^k(\tilde{t}_{i+1}) + \tilde{y}^k(\tilde{t}_i))'g(\tilde{t}_i+) - \\
&= \frac{\Delta t_i + \Delta t}{2}(\tilde{y}^k(\tilde{t}_i) + \tilde{y}^k(\tilde{t}_{i-1}) + \Delta t y_1)'g(\tilde{t}_{i-1}+) + \\
&\quad \frac{\Delta t_{i+1} - \Delta t}{2}(\tilde{y}^k(\tilde{t}_{i+1}) + \tilde{y}^k(\tilde{t}_i) + \Delta t y_1)'g(\tilde{t}_i+) - \\
&\quad \frac{\Delta t_i}{2}(\tilde{y}^k(\tilde{t}_i) + \tilde{y}^k(\tilde{t}_{i-1}))'g(\tilde{t}_{i-1}+) - \frac{\Delta t_{i+1}}{2}(\tilde{y}^k(\tilde{t}_{i+1}) + \tilde{y}^k(\tilde{t}_i))'g(\tilde{t}_i+) \\
&= \frac{\Delta t}{2}(\tilde{y}^k(\tilde{t}_{i-1}) - \tilde{y}^k(\tilde{t}_{i+1}) + (\tilde{t}_{i+1} - \tilde{t}_{i-1})y_1)'g(t_l+). \qquad (3.32)
\end{aligned}
$$

Combining (3.31) and (3.32), we derive

$$
\begin{aligned}
&\Psi(\tilde{v}^*,\ \tilde{y}^*,\ \tilde{t}^*) - \Psi(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k) \\
&= \frac{\Delta t \Delta t_i}{2}(\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i-) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_i) - \tilde{y}^k(\tilde{t}_{i-1})}{\Delta t_i} - (\dot{c}(\tilde{t}_i)'u_1^{(i)} - y_1'g(t_l+))) + \\
&\quad \frac{\Delta t \Delta t_{i+1}}{2}(\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} - (\dot{c}(\tilde{t}_i)'u_1^{(i)} - y_1'g(t_l+))) \\
&\qquad (3.33)
\end{aligned}
$$

By the definition of a KKT point and the discussion following it in Section 3.3, a feasible solution to $QP(|P|)$ is a KKT point if and only if there is no feasible descent direction for this solution. Hence

$$\Psi(\tilde{v}^*,\ \tilde{y}^*,\ \tilde{t}^*) - \Psi(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k) \geq 0. \tag{3.34}$$

Thus, (3.33) implies that $\dot{c}(\tilde{t}_i)'u_1^{(i)} - y_1'g(t_l+)$ is uniformly bounded from above.

For any $\bar{j} \leq q(i)$ and any $\epsilon \in (0\ ,1)$, we have

$$\begin{aligned}\tilde{u}^k(\tilde{t}_i+) &= \epsilon\lambda_1^{(i)}(s_1^{(i)} + \frac{\mu_{\bar{j}}^{(i)}}{\epsilon\lambda_1^{(i)}}r_{\bar{j}}^{(i)}) + \\ &\quad (1-\epsilon\lambda_1^{(i)})\left(\sum_{j=2}^{k(i)}\frac{\lambda_j^{(i)}}{1-\epsilon\lambda_1^{(i)}}s_j^{(i)} + \frac{\lambda_1^{(i)}(1-\epsilon)}{1-\epsilon\lambda_1^{(i)}}s_1^{(1)} + \sum_{j=1,j\neq\bar{j}}^{q(i)}\frac{\mu_j^{(i)}}{1-\epsilon\lambda_1^{(i)}}r_j^{(i)}\right)\end{aligned}$$

and

$$\begin{aligned}&\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} \\ &= \epsilon\lambda_1^{(i)}(\bar{s}_1^{(i)} + \frac{\mu_{\bar{j}}^{(i)}}{\epsilon\lambda_1^{(i)}}\bar{r}_{\bar{j}}^{(i)}) + \\ &\quad (1-\epsilon\lambda_1^{(i)})\left(\sum_{j=2}^{k(i)}\frac{\lambda_j^{(i)}}{1-\epsilon\lambda_1^{(i)}}\bar{s}_j^{(i)} + \frac{\lambda_1^{(i)}(1-\epsilon)}{1-\epsilon\lambda_1^{(i)}}\bar{s}_1^{(1)} + \sum_{j=1,j\neq\bar{j}}^{q(i)}\frac{\mu_j^{(i)}}{1-\epsilon\lambda_1^{(i)}}\bar{r}_j^{(i)}\right).\end{aligned}$$

By letting $u_1 = s_1^{(i)} + \frac{\mu_{\bar{j}}^{(i)}}{\epsilon\lambda_1^{(i)}}r_{\bar{j}}^{(i)}$, $y_1 = \bar{s}_1^{(i)} + \frac{\mu_{\bar{j}}^{(i)}}{\epsilon\lambda_1^{(i)}}\bar{r}_{\bar{j}}^{(i)}$ and letting ϵ tend to zero, the above boundedness result on $\dot{c}(\tilde{t}_i)'u_1^{(i)} - y_1'g(t_l+)$ implies (3.30). Since $[t_{i-1},\ t_i]$ is not the first positive length interval that resides on $[t_l,\ t_m]$, we can similarly have

$$\dot{c}(\tilde{t}_i)'r_j^{(i-1)} - g(t_l+)'\bar{r}_j^{(i-1)} \leq 0, \quad \text{for all } j.$$

These together with (3.28) and (3.29) give

$$\dot{c}(\tilde{t}_i)'s_1^{(i-1)} - g(t_l+)'\bar{s}_1^{(i-1)} \geq \dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_{i-1}+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_i) - \tilde{y}^k(\tilde{t}_{i-1})}{\Delta t_i} \tag{3.35}$$

Similarly,

$$\dot{c}(\tilde{t}_i)'s_1^{(i)} - g(t_l+)'\bar{s}_1^{(i)} \geq \dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}}. \tag{3.36}$$

Since $\tilde{P}$ is a purified partition, by Procedure $\mathcal{A}$, the opposite of (3.9) holds which is equivalent to

$$\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_{i-1}+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_i) - \tilde{y}^k(\tilde{t}_{i-1})}{\Delta t_i} > \dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} \tag{3.37}$$

Now, suppose

$$\dot{c}(\tilde{t}_i)'s_1^{(i-1)} - g(t_l+)'\bar{s}_1^{(i-1)} \leq \dot{c}(\tilde{t}_i)'s_1^{(i)} - g(t_l+)'\bar{s}_1^{(i)}.$$

By (3.35) and (3.37), we have

$$\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_i+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_{i+1}) - \tilde{y}^k(\tilde{t}_i)}{\Delta t_{i+1}} < \dot{c}(\tilde{t}_i)'s_1^{(i)} - g(t_l+)'\bar{s}_1^{(i)}$$

and

$$\dot{c}(\tilde{t}_i)'\tilde{u}^k(\tilde{t}_{i-1}+) - g(\tilde{t}_i+)'\frac{\tilde{y}^k(\tilde{t}_i) - \tilde{y}^k(\tilde{t}_{i-1})}{\Delta t_i} \leq \dot{c}(\tilde{t}_i)'s_1^{(i)} - g(t_l+)'\bar{s}_1^{(i)}.$$

Let $u_1 = s_1^{(i)}$ and $y_1 = \bar{s}_1^{(i)}$, and the above relationship together with (3.33) gives

$$\Psi(\tilde{v}^*,\ \tilde{y}^*,\ \tilde{t}^*) - \Psi(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k) < 0,$$

which contradicts the fact that $(\tilde{v}^k,\ \tilde{y}^k,\ \tilde{t}^k)$ is a KKT point for $QP(|\tilde{P}^k|)$ (cf. (3.34)). □

Since $\dot{c}(\tilde{t}_i)$ is a constant vector over $[t_l,\ t_m]$, as a consequence of Lemma 3.7, every nonzero length interval that resides on $[t_l,\ t_m]$ (except the first nonzero length interval) corresponds to a different generalized extreme point of some system (SYS_{J_i}). Since only a finite number of different systems (SYS_{J_i}) exist, and for each (SYS_{J_i}) there are a finite number of generalized extreme points, we see there are only a finite number of nonzero length intervals that reside on $[t_l,\ t_m]$. Since the number of zero length intervals that reside on $[t_l,\ t_m]$ is at most 2 (one on each end of $[t_l,\ t_m]$), there are also a finite number of breakpoints in $[t_l, \ t_m]$. Thus we have the following corollary.

Corollary 3.3 *There are a finite number of breakpoints in $\tilde{P}^k$.*

There are only a finite number of different solution values for all the KKT points of $QP(|P|)$, as shown in the following lemma.

Lemma 3.8 *The KKT points for $QP(|P|)$ is the union of a finite number of connected sets. Over each connected component of KKT points of $QP(|P|)$, the objective value is a constant. Furthermore, the number of connected sets is bounded from above by a number that depends on $|P|$ only.*

Proof It is easily seen that a solution to $QP(|P|)$ is a KKT point of $QP(|P|)$ if and only if it is a solution to a feasible symmetric affine variational inequality problem whose dimension depends only on $|P|$ (cf. Section 3.3). The lemma now follows directly from Lemma 3.1 of Luo and Tseng [62]. □

We now present the main convergence result of this chapter.

Theorem 3.7 *Algorithm $\mathcal{A}$ will terminate after a finite number of iterations.*

Proof Suppose Algorithm $\mathcal{A}$ does not terminate after finite number of iterations. It is guaranteed by Theorem 3.5 that step 4 of Algorithm $\mathcal{A}$ would produce a strictly improved solution, and thus every iteration of Algorithm $\mathcal{A}$ would give a KKT point of certain $QP(|P|)$ that has strictly better solution value. By Lemma 3.8, the KKT points generated by $QP(|P|)$ should lie on a different connected KKT points component of $QP(|P|)$ for every $|P|$. This means that the cardinality of P is unbounded and contradicts Corollary 3.3. □

3.8 New Structural and Duality Results

As a result of Algorithm $\mathcal{A}$ and Theorem 3.7, we have the following new structural result for $(SCSCLP)$.

Theorem 3.8 *Algorithm $\mathcal{A}$ terminates with a solution to $QP(|P|)$ for some P that gives the optimal objective value of $(SCSCLP)$ and can be closely approximated by* *a series of piecewise constant controls for* $(SCSCLP)$. *When the solution set for* $(SCSCLP)$ *is bounded and* E *is an identity matrix, Algorithm* $\mathcal{A}$ *terminates with a piecewise constant optimal control with partition* P *such that* $t_i \neq t_{i-1}$ *for all* i. *Furthermore, over each interval* $[t_{i-1},\ t_i)$, $(u(t_i+), \frac{y(t_{i+1})-y(t_i)}{t_{i+1}-t_i})$ *is a convex combination of the generalized extreme points of linear system* (SYS_{J_i}), *where* J_i *is a subset of* $\{1, \cdots, n_2\}$.

Proof The first part of the theorem is a direct consequence of Theorem 3.7. The second part of the theorem follows from Lemma 3.5 and the remark following the proof of Lemma 3.1. □

We also derive the following new duality result for $(SCSCLP)$.

Theorem 3.9 *There is no duality gap between* $(SCSCLP)$ *and* $(SCSCLP^*)$. *There always exists an optimal solution for* $(SCSCLP^*)$ *that is piecewise linear. Furthermore, there exists a bounded measurable optimal solution for* $(SCSCLP)$ *if and only if Algorithm* $\mathcal{A}$ *terminates with such a solution.*

Proof The first part of the theorem is a direct consequence of Theorem 3.7.

Denote $\tilde{P}^k$ as the final purified partition when Algorithm $\mathcal{A}$ terminates. To prove the second part of the theorem, we first show that the zero length interval in $\tilde{P}^k$ can be eliminated in the dual problem $AP^*(P)$. Let $[t_{i-1},\ t_i]$ be a zero length interval that resides on $[t_l,\ t_m]$ where t_l and t_m are two consecutive breakpoints in $D_1^{\tilde{P}^k}$. By Lemma 3.4, the zero length intervals can be located only at the breakpoints in $D_1^{\tilde{P}^k}$. We assume $t_i = t_m$ (the case $t_{i-1} = t_l$ can be treated similarly).

Let $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ be an optimal solution for $AP^*(\tilde{P}^k)$ and we construct a new solution $(\tilde{\pi},\ \tilde{\eta},\ \tilde{\xi})$ for $AP^*(\tilde{P}^k)$ in the following way. Let $(\tilde{\pi},\ \tilde{\eta},\ \tilde{\xi})$ equal $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ except

$$\begin{aligned} \tilde{\pi}(t_{i-1}+) &= \hat{\pi}(t_{i-1}-), \quad \tilde{\pi}(t_i-) = \hat{\pi}(t_{i-1}-) \\ \tilde{\eta}(t_{i-1}+) &= \hat{\eta}(t_{i-1}-), \quad \tilde{\eta}(t_i-) = \hat{\eta}(t_{i-1}-) \\ \tilde{\xi}\left(\frac{t_{i-1}+t_i}{2}\right) &= 0, \qquad \tilde{\xi}(t_{i-1}) = 0 \\ \tilde{\xi}(t_i) &= \hat{\xi}\left(\frac{t_{i-1}+t_i}{2}\right) + \hat{\xi}(t_{i-1}) + \hat{\xi}(t_i) \end{aligned}$$

It is easy to check the feasibility of $(\tilde{\pi},\ \tilde{\eta},\ \tilde{\xi})$. It is a fact that $(\tilde{\pi},\ \tilde{\eta},\ \tilde{\xi})$ and $(\hat{\pi},\ \hat{\eta},\ \hat{\xi})$ have the same solution value in $AP^*(\tilde{P}^k)$. Let $\bar{P}$ be $\tilde{P}^k \setminus \{t_{i-1}\}$. By eliminating the elements $\tilde{\pi}(t_{i-1}-)$, $\tilde{\pi}(t_{i-1}+)$, $\tilde{\eta}(t_{i-1}-)$, $\tilde{\eta}(t_{i-1}+)$, $\tilde{\xi}(t_{i-1})$ and $\tilde{\xi}\left(\frac{t_{i-1}+t_i}{2}\right)$ from $(\tilde{\pi},\ \tilde{\eta},\ \tilde{\xi})$, we can get a feasible solution $(\bar{\pi},\ \bar{\eta},\ \bar{\xi})$ for $AP(\bar{P})$. Also, $(\bar{\pi},\ \bar{\eta},\ \bar{\xi})$ has the same solution value as $(\tilde{\pi},\ \tilde{\eta},\ \tilde{\xi})$.

By repeating this process, we can eliminate all the zero length intervals in $\tilde{P}^k$ and define a feasible solution for $AP^*(P)$ from the resulting partition P. From this feasible solution we can construct an optimal solution for $(SCSCLP^*)$ that is piecewise linear. This proves the second part of the theorem.

One direction of the third part of the theorem is quite obvious. The other direction (i.e., if there exists a bounded measurable optimal solution for $(SCSCLP)$, then Algorithm $\mathcal{A}$ will find such a solution) can be shown as follows. Let the bounded measurable solution $(u(t),\ y(t))$ be optimal for $(SCSCLP)$. By the second part of the theorem, there always exists an optimal solution $(\pi(t),\ \eta(t),\ \xi(t))$ for $(SCSCLP^*)$ that is piecewise linear with partition P (defined by removing all the zero length intervals from $\tilde{P}^K$). By Corollary 2.1, the complementary slackness condition (2.1) is satisfied. Let $\bar{u}(t)$ be the piecewise constant extensions of $u(t_0+)$, $u(t_1+)$, $\cdots$, $u(t_{p-1}+)$. Let $\bar{y}(t)$ be the piecewise linear extension of $y(t_0+), y(t_1-)\ y(t_1+),\ \cdots,\ y(t_{p-1}+), y(t_p-)$. The solution $(\bar{u}(t),\ \bar{y}(t))$ is a feasible solution for $(SCSCLP)$ which together with $(\pi(t),\ \eta(t),\ \xi(t))$ satisfies (2.1). By Corollary 2.1 again, $(\bar{u}(t),\ \bar{y}(t))$ is optimal for $(SCSCLP)$. □

Chapter 4

Applications of Fluid Networks

Stochastic optimal control of queueing networks has many important applications in communication and manufacturing. These problems are however, very difficult to solve. Following the approach by Anderson [6], Hajek and Ogier [42], Chen and Yao [20], this chapter analyzes the fluid flow approximation of several queueing control problems, which include the single multiclass queueing control problem and the single class tandem queueing control problem. We apply the theory of continuous linear programming to these queueing control problems. We either give polynomial time algorithms to solve the problem or indicate strong evidence about the existence of such algorithms for the problem. We demonstrate strong ties between the solution of fluid flow approximation and the optimal solution to the stochastic queueing control models. We also investigate conditions under which a general class of fluid networks is stabilizable. The chapter is structured as follows. In Section 4.1, we introduce the concept of fluid networks. In Section 4.2, we analyze the fluid approximation of a single multiclass queue first introduced by Klimov [54]. In Section 4.3, we consider a single multiclass queue with convex separable quadratic cost. In Section 4.4, we analyze the fluid approximation of the tandem queueing network and its variations. In Section 4.5, we give simple necessary and sufficient conditions for the fluid networks to be stabilizable.

4.1 The Linear Fluid Networks

A *Linear fluid network* consists of a network of m stations connected by n nonempty finite sets of links (routes), which we denote by L_i, for $i = 1, \cdots, n$. We let $L = \bigcup_i L_i$ and assume $L_i \cap L_j = \emptyset$ for all $i \neq j$. There are n classes of customers (fluids, inventory or traffic), one for every set L_i. Customers of class i can use only the links in set L_i and can be served only at a specified station, which we denote by $\mathcal{S}(i)$. We denote the set of all customer classes that are served at station j as $\mathcal{C}(j)$. We let b_i be the rate at which class i customers arrive at or depart from station $\mathcal{S}(i)$ depending on whether b_i is nonnegative. Among the customers going through link $r \in L_i$, a fraction p_{jr} of them become class j customers and go to station $\mathcal{S}(j)$, and the remaining fraction $1 - \sum_j p_{jr}$ exit the network. A station can work on an arbitrary number of links in L_i simultaneously. For $r \in L_i$, we let variable $u_r(t)$ be the rate at which class i customers go through link r. We require these rates satisfy $\tilde{n}$ capacity constraints $Du(t) \leq c$, where D is a $\tilde{n} \times n$ nonnegative matrix with no zero columns and c is a $\tilde{n} \times 1$ vector. Our objective is to find an optimal control policy (involving both routing and sequencing decisions) that minimizes the cumulated cost of queueing over a fixed time horizon $[0,\ T]$. We will also investigate the existence of controls that can eventually drive all the queues in the network to zero or make the total queue length stay bounded for bounded initial conditions. We require that the queue length for each class of customers stay nonnegative. We will also discuss extensions of our results for the situations permitting backloging.

Let variables $x_i(t)$, $i = 1, \cdots, n$ be the queue length of class i customers at time t. We can formulate the problem as follows:

$$
\begin{aligned}
(FNET) \quad \textit{minimize} \quad & \int_0^T w'x(t)\, dt \\
\textit{subject to} \quad & x(t) = x(0) + \int_0^t (Bu(t) + b)\, dt && (4.1) \\
& Du(t) \leq c && (4.2) \\
& x(t) \geq 0,\ u(t) \geq 0 \qquad t \in [0,\ T],
\end{aligned}
$$

where $x(0) \geq 0$, b is a given vector, w is a given weight vector and B is an $n \times |L|$

matrix with

$$B_{ir} = \begin{cases} p_{ir}, & \text{if } r \notin L_i \\ p_{ir} - 1, & \text{if } r \in L_i. \end{cases}$$

Note that $(FNET)$ includes as a special case the fluid approximation of a Generalized Jackson Network (see Jackson [49]) when $|L_i| = 1$ for all i and D is an identity matrix. It includes as a special case the continuous traffic control problem in a communication network considered by Hajek and Ogier [42] when p_{ir} are either 0 or 1 and D is an identity matrix. We can also show that the fluid networks include the fluid approximation for multiclass queueing networks considered in Chen and Yao [20], when $|L_i| = 1$, b is the arrival rate, c is the vector of all ones and D is a block diagonal matrix, with each block a row vector of mean service times of the customers served at the same workstation. The re-entrant line considered in Kumar [56] is a multiclass queueing network with fixed routing, so $(FNET)$ includes as a special case the re-entrant lines as well. Chen and Mandelbaum [19], Chen [21], Dai [27] and Chen [22] showed that a wide range of queueing networks would converge (under appropriate time and space scaling) to fluid networks, providing theoretical justifications for using fluid model as approximations for queueing networks. We note that $(FNET)$ is an SCLP.

4.2 The Klimov's Problem

Consider the optimal control for the fluid flow approximation to the multiclass queueing scheduling problem first introduced by Klimov [54]. As a queueing control problem, this problem can be modeled as a multi-armed bandit problem, and the optimal control is known to be a priority index rule. The indices can be calculated through a dual algorithm (see Bertsimas and Nino-Mora [15]). Chen and Yao in [20] considered a myopic solution procedure for the fluid network. They showed that their algorithm will produce the priority index rule that solves the Klimov's problem (under the long term average cost objective). However, to establish the optimality of their indices for the fluid network, additional conditions had to be imposed (see [20] [Theorem. 4.1]). In this section, we propose a control policy for the fluid Klimov's problem. We show that the policy is optimal for the fluid network and in addition, the policy produces the optimal index rule for the Klimov's problem as well.

4.2.1 The Fluid Klimov's Problem

The problem we consider is the following. There are n classes of customers arriving at a single queue. After finishing service, a fraction p_{ij} of class i customers change to class j customers and the remaining fraction $1 - \sum_j p_{ij}$ exit the system. The arrival rate of class i customers is b_i and the service rate of class i customers is $\frac{1}{a_i}$. The cost per unit time for a class i customer waiting in the queue is w_i. The service facility has to dynamically decide which job class, if any, to serve next in order to minimize the average cost incurred per unit of time over the interval $[0,\ T]$, where the costs are the waiting costs for the customers stuck in the queue. The problem is shown in Figure 4-1.

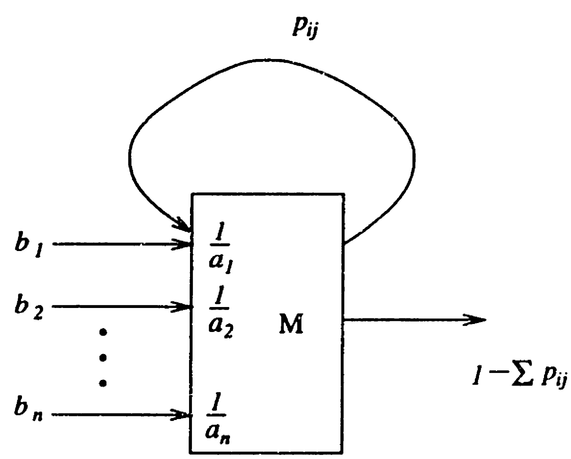

Figure 4-1: The Klimov's problem

The problem can be formulated as follows:

$$(MBLP) \quad minimize \quad \int_0^T w'x(t)\, dt$$

$$\text{subject to} \quad x(t) = x(0) + \int_0^t (Bu(t) + b)\, dt \tag{4.3}$$
$$a'u(t) \le 1 \tag{4.4}$$
$$x(t) \ge 0, \quad u(t) \ge 0,$$

where $a > 0$, $b \ge 0$, $x(0)$ and $T \ge 0$ are given and fixed, $B = A - I$, A is a substochastic matrix, i.e., a square matrix with nonnegative entries and whose column sums are less than or equal to ones. Here, I is an identity matrix.

By introducing dual variables $\pi(t)$, $\eta(t)$ for constraints (4.3) and (4.4) respectively, we have the following dual problem

$$(MBLP^*) \quad \text{maximize} \quad \int_0^T (x(0) + bt)'\pi(t)\, dt - \int_0^T \eta(t)\, dt$$
$$\text{subject to} \quad \int_t^T B'\pi(t)\, dt + a\eta(t) \ge 0 \tag{4.5}$$
$$\pi(t) \le w \tag{4.6}$$
$$\eta(t) \ge 0.$$

As a consequence of Proposition 2.4, we have

Corollary 4.1 *Weak duality holds between $(MBLP)$ and $(MBLP^*)$. Moreover, if*

$$x(t)'(w - \pi(t)) = 0 \tag{4.7}$$
$$u(t)'(a\eta(t) + \int_t^T B'\pi(t)\, dt) = 0 \quad \text{for all } t \tag{4.8}$$
$$\eta(t)(a'u(t) - 1) = 0 \quad \text{for all } t \tag{4.9}$$

hold, then strong duality holds between $(MBLP)$ and $(MBLP^*)$.

Relations (4.7)–(4.9) are the complementary slackness conditions for $(MBLP)$ and $(MBLP^*)$.

4.2.2 Properties of Matrix B

Before proposing a control policy, we first give some properties of the matrix B. Some of the results in this section can be found in standard linear algebra textbooks, such as Seneta [85]. For completeness of the thesis, we include their proof.

For any n dimensional vector x, we denote by x_i the i-th coordinate of x, and, for any nonempty subset $Q \subseteq \{1, \cdots, n\}$, we use x_Q and $[x]_Q$ to denote the the vector with components x_i, $i \in Q$ (with x_i arranged in the same order as in x).

Lemma 4.1 *Let M be an $n \times m$ matrix, $1 \leq m \leq n$. Assume that all the off-diagonal elements of M are nonnegative, M is columnwise diagonally dominant in the following sense:*

$$-M_{jj} \geq \sum_{i \neq j} M_{ij} \quad \forall j \leq m.$$

We have
1) If M is of full column-rank, then the submatrix formed by eliminating rows from $m+1$ to n is a full rank square matrix.
2) If M is not of full column-rank then there exists $R \neq \emptyset$, $R \subseteq \{1, 2, \cdots, m\}$ such that

$$\sum_{i \in R} M_{ij} = 0 \quad \forall j \in R,$$

and

$$\sum_{i \notin R} M_{ij} = 0 \quad \forall j \in R.$$

Proof We prove the first part of the lemma by induction on the dimension m. If $m = 1$, the result trivially holds. Suppose the result holds for $m = k \leq n-1$. We will show the result holds for $m = k+1$. If $\sum_{i \geq k+2} M_{ij} = 0$ for all $j \leq m$, we are all set. Suppose $\sum_{i \geq k+2} M_{ij} > 0$ for some $j \leq m$, we let $\tilde{M}_{il} = M_{il} - \frac{M_{jl}}{M_{jj}} M_{ij}$ for all $i \leq n$ and $l \neq j$ and let $\tilde{M}_{ij} = M_{ij}$ for all $i \leq n$. The submatrix $\bar{M}$ of $\tilde{M}$ formed by eliminating from $\tilde{M}$ column j and row j satisfies the induction hypothesis, so the submatrix by eliminating from $\bar{M}$ rows from $k+1$ to $n-1$ is a full rank square matrix, so is the matrix by eliminating from $\tilde{M}$ rows from $k+2$ to n. Thus the submatrix by eliminating from M the rows from $k+2$ to n is a full rank square matrix. This completes the induction for the first part of the lemma.

We now prove the second part of the lemma. Since M is not of full column-rank, the first m rows of M must be linearly dependent. So there exist α_1, α_2, $\cdots$, α_m which are not identically zero such that

$$\sum_{i \leq m} \alpha_i M_{ij} = 0 \qquad \forall j \tag{4.10}$$

Let $R = \{ i \mid \alpha_i = \max_k \alpha_k\}$ and $S = \{ i \mid \alpha_i = \min_k \alpha_k\}$. If $\min_k \alpha_k \geq 0$ then $\max_k \alpha_k > 0$. From (4.10) we see $\sum_{i\in R} M_{ij} = 0$, for all $j \in R$ and $M_{ij} = 0$ for all $i \notin R$ and $j \in R$. If $\min_k \alpha_k < 0$, from (4.10) we see $\sum_{i\in S} M_{ij} = 0$, for all $j \in S$ and $M_{ij} = 0$ for all $i \notin S$ and $j \in S$. Obviously, $R \neq \emptyset$, $S \neq \emptyset$, $R \subseteq \{1, 2, \cdots, m\}$ and $S \subseteq \{1, 2, \cdots, m\}$ and therefore the result follows. □

Lemma 4.2 *a) The determinant of B is nonpositive.*

b) The adjoint matrix of B has non-positive entries.

c) Furthermore, if the system $Bx = b$ has a solution then it must have a nonpositive solution given that b is a nonnegative vector.

The same results hold for B'.

Proof The nonpositiveness of the determinant of B and B' can be proven by applying standard Gaussian eliminating procedure to the matrix B', since the procedure does not change the diagonal dominance and the nonnegativeness of the off-diagonal elements of the matrix. The B matrix is invertible if it is strictly dominated by the diagonal elements.

We first prove that the following system has a solution for every $\epsilon > 0$:

$$Bx \leq \epsilon x, \; x \geq e, \tag{4.11}$$

where e is the vector of all ones. Consider the linear program

$$\begin{array}{rl} \max & e'z \\ \text{such that} & B'y - \epsilon y + z = 0 \\ & y \leq 0, \; z \geq 0 \end{array} \tag{4.12}$$

which is the dual problem to

$$\min \; 0 \text{ s. t. } \; Bx \leq \epsilon x, \; x \geq e$$

The dual problem is feasible ($y = 0$, $z = 0$ is feasible). For any feasible solution (y, z) to the dual problem, let y_j be the smallest element of y. Since the column sums of B are less than or equal to zero, the j–th equality of (4.12) implies $y_j = 0$, otherwise the left hand side of the j–th equality would be strictly positive. So $y = 0$ and $z = 0$ is the only feasible solution to the dual problem. So the dual problem has a finite optimal solution value. By linear programming duality theory, we see the primal problem is feasible.

Let $B_\epsilon = B - \epsilon I$. Suppose x_ϵ is a feasible solution to (4.11). Let $X = \text{diag}(x_\epsilon)$. It is easy to see the columns of $B_\epsilon X$ are strictly dominated by the diagonal elements. By (4.11), its rows are dominated by the diagonal elements. So $B_\epsilon X$ is a negative definite matrix. In order to prove the adjoint matrix of B has non-positive entries, it suffices to prove the adjoint matrix of B_ϵ has non-positive entries for all $\epsilon > 0$ due to continuity. So it suffices to prove $B_\epsilon^{-1} b \leq 0$ for all $b \geq 0$ (since in particular it is true when b is chosen as a unit vector, which means the corresponding column of B_ϵ is non-positive). Suppose $B_\epsilon^{-1} b_\epsilon \leq 0$ is infeasible for some $b_\epsilon \geq 0$. The dual of

$$\min \quad 0 \;\; \text{s. t.} \quad B_\epsilon x = b_\epsilon, \; x \leq 0$$

is

$$\begin{aligned} \max \quad & b_\epsilon' y \\ \text{such that} \quad & B_\epsilon' y \geq 0 \end{aligned} \tag{4.13}$$

The dual is obviously feasible. By linear programming duality, there is a dual solution y with $b_\epsilon y > 0$. Let y_P be the maximal subvector of y such that $y_P > 0$. We denote the remaining of y as y_Q. It is easy to see $y_Q \leq 0$. Multiplying the i–th constraint in (4.13) by $X_{ii} y_i$ and summing up those corresponding to positive y_is, we have

$$y_P' B_{PP} X_{PP} y_P - \epsilon y_P' X_{PP} y_P + y_Q' B_{QP} X_{PP} y_P \geq 0,$$

which is impossible in view of the negative definiteness of $B_{PP} X_{PP} - \epsilon X_{PP}$, the positivity of y_P and the nonnegativeness of B_{QP} and $-y_Q$. So the adjoint matrix of B has non-positive entries. The adjoint matrix of B' is the transpose of the adjoint matrix of B and thus also has nonpositive entries.

Now we prove the final property of the lemma. We prove it constructively. Let P be the index set to a set of maximal independent columns of B. From linear algebra, the assumption $Bx = b$ is feasible implies $B_{\bullet P}x_P = b$ is feasible, where $B_{\bullet P}$ is the submatrix of B by removing all the columns not indexed by P. From Lemma 4.1, we see $B_{\bullet P}x_P = b$ if and only if

$$B_{PP}x_P = b_P \tag{4.14}$$

By the same arguments as in the previous paragraph, we see the matrix B_{PP}^{-1} has non-positive entries, which implies that the solution to (4.14) satisfies $x_P \leq 0$. By setting the remaining elements of x to zero, we have a feasible solution to $Bx = b,\ x \leq 0$.

Next, we prove this property for B'. Since the problem

$$\min \quad 0 \quad \text{s. t.} \quad B'x = b$$

is feasible, we see the dual problem

$$\max \quad b'y \quad \text{s. t.} \quad By = 0$$

is feasible and has finite optimal solution value. Suppose

$$\min \quad 0 \quad \text{s. t.} \quad B'x = b,\ x \leq 0$$

is infeasible, by linear programming duality, there exists y_1 such that

$$b'y_1 > 0 \quad \text{and} \qquad By_1 \geq 0.$$

By the property of B proved in the previous paragraph, there exists $y_2 \leq 0$ such that $By_2 = By_1$. So

$$b'(y_1 - y_2) > 0 \quad \text{and} \qquad B(y_1 - y_2) = 0,$$

contradicting the fact that the problem

$$\max \quad b'y \quad \text{s. t.} \quad By = 0$$

is feasible and has finite optimal solution value. □

We let $N = \{1, \cdots, n\}$, the set of all customer classes.

Lemma 4.3 *For any two subsets P and $Q = N \setminus P$ of N, the following system always has a solution*

$$[B'x]_P = 0, \qquad x_Q = w_Q. \tag{4.15}$$

Proof We prove the lemma by using induction on the cardinality of the set P. The lemma trivially holds if P is either empty or a singleton. Suppose the lemma holds for $|P| \leq k \leq n-1$. We will show the lemma holds for $|P| = k+1$. Let $B_{\bullet P}$ be the submatrix of B by removing all the columns not indexed by P. If $B_{\bullet P}$ is of full column-rank, the result follows easily. Suppose $B_{\bullet P}$ is not of full column-rank. By the second property of Lemma 4.1, there exists $R \neq \emptyset$, $R \subseteq P$ such that

$$\sum_{i \in R} B_{ij} = 0 \quad \forall j \in R,$$

and

$$\sum_{i \in N \setminus R} B_{ij} = 0 \quad \forall j \in R.$$

It is easy to see that $[B'x]_R = 0$ for all x such that $x_R = 0$. Consider the following problem,

$$[B'x]_{P \setminus R} = 0, \qquad x_Q = w_Q, \qquad x_R = 0 \tag{4.16}$$

By induction hypothesis, we know (4.16) is feasible. The same x would be feasible for (4.15).

□

Lemma 4.4 *Let $R \subset N$, $r_1 \in N$ and $r_2 \in N$ such that $r_1 \neq r_2$. $r_1 \notin R$ and $r_2 \notin R$. Let $P = R \cup \{r_1\}$ and $Q = R \cup \{r_1\} \cup \{r_2\}$. Suppose there exists u^* such that*

$$\begin{aligned} [Bu^* + b]_R &= 0 \\ [Bu^* + b]_{r_1} &\leq 0 \\ a'u^* &= 1 \\ u^*_P \geq 0, \quad u^*_{N \setminus P} &= 0, \end{aligned} \tag{4.17}$$

then there exist $\tilde{u}$ and $\tilde{u}$ such that

$$\begin{aligned} [B\tilde{u} + b]_P &= 0 \\ a'\tilde{u} &= 1 \end{aligned}$$

$$\tilde{u}_Q \geq 0, \quad \tilde{u}_{N\backslash Q} = 0 \tag{4.18}$$

and

$$\begin{aligned} [B\bar{u} + b]_P &= 0 \\ a'\bar{u} &\leq 1 \\ \bar{u}_P \geq 0, \quad \bar{u}_{N\backslash P} &= 0. \end{aligned} \tag{4.19}$$

Proof If $[Bu^* + b]_{r_1} = 0$, we may choose $\tilde{u} = u^*$ and $\bar{u} = u^*$. Suppose $[Bu^* + b]_{r_1} = \delta < 0$. For any $\epsilon > 0$, we denote $B_\epsilon = B - \epsilon I$. From system (4.17), we see

$$\begin{aligned} [B_\epsilon u^* + b]_R &= -\epsilon u^*_R \\ [B_\epsilon u^* + b]_{r_1} &= -\epsilon u^*_{r_1} + \delta \\ a'u^* &= 1 \\ u^*_P \geq 0, \quad u^*_{N\backslash P} &= 0. \end{aligned} \tag{4.20}$$

We first prove the following system has a solution:

$$[B_\epsilon x]_P = \begin{pmatrix} 0 \\ -\delta \end{pmatrix} \tag{4.21}$$

$$a'x = 0 \tag{4.22}$$

$$x_{r_2} \geq 0, \quad x_{N\backslash Q} = 0$$

We use the following argument. Let $M_1 = B_{PP} - \epsilon I$ and let l_2 be the subvector of the r_2–th column of B indexed by P. By Lemma 4.2, M_1^{-1} is nonpositive. Given any x such that $x_{N\backslash Q} = 0$, (4.21) implies

$$x_P = M_1^{-1} \begin{pmatrix} 0 \\ -\delta \end{pmatrix} - x_{r_2} M_1^{-1} l_2.$$

So

$$a'x = a'_P M_1^{-1} \begin{pmatrix} 0 \\ -\delta \end{pmatrix} - x_{r_2} a'_P M_1^{-1} l_2 + a_{r_2} x_{r_2}. \tag{4.23}$$

If $x_{r_2} = 0$, $a'x < 0$. For x_{r_2} large enough, $a'x > 0$. Since the mapping in the right hand side of (4.23) is continuous, there exists a $x^*_{r_2} > 0$, together with $x^*_{N\backslash Q} = 0$ and the unique x^*_P specified by (4.21) is feasible for the system. Let $y^*_P = M_1^{-1}\begin{pmatrix} 0 \\ -\delta \end{pmatrix}$ and $y^*_{N\backslash P} = 0$. It is easy to see that y^* is a feasible solution to the following system:

$$\begin{aligned} [B_\epsilon y]_P &= \begin{pmatrix} 0 \\ -\delta \end{pmatrix} \\ a'y &\leq 0 \\ y_{r_1} \leq 0, \quad y_{N\backslash P} &= 0. \end{aligned}$$

Now, let $\tilde{u}_\epsilon = u^* + x^*$. It is certainly feasible for the following system:

$$\begin{aligned} [B_\epsilon u + b]_R &= -\epsilon u^*_R & (4.24) \\ [B_\epsilon u + b]_{r_1} &= -\epsilon u^*_{r_1} & (4.25) \\ a'u &= 1 \\ u_{r_2} \geq 0, \quad u_{N\backslash Q} &= 0. \end{aligned}$$

Since $\tilde{u}_{\epsilon r_2} \geq 0$, by Lemma 4.2, (4.24) and (4.25) ensure the unique vector $\tilde{u}_{\epsilon P}$ is nonnegative. Thus $\tilde{u}_\epsilon$ is also feasible for the following system:

$$\begin{aligned} [B_\epsilon u + b]_R &= -\epsilon u^*_R \\ [B_\epsilon u + b]_{r_1} &= -\epsilon u^*_{r_1} \\ a'u &= 1 \\ u_Q \geq 0, \quad u_{N\backslash Q} &= 0. \end{aligned}$$

Given any sequence $\{\epsilon_i\}$, $\{\tilde{u}_{\epsilon_i}\}$ belongs to a compact set, since $a'\tilde{u}_{\epsilon_i} \leq 1$. So there always exists a cluster point for $\{\tilde{u}_{\epsilon_i}\}$. If in addition

$$\lim_{k\to\infty} \epsilon_k = 0$$

and $\tilde{u}$ is a cluster point of $\{\tilde{u}_{\epsilon_i}\}$, we see $\tilde{u}$ is a feasible solution for system (4.18). Similarly, we can prove $\bar{u} = u^* + y^*$ is a feasible solution for system (4.19). □

If we let R be a set of customer classes and let r_1 and r_2 be two different customer classes that do not belong to R, assuming that we have a control that uses the full capacity, it only works on $R \cup \{r_1\}$ classes of customers, the customers in R are maintained at zero level, and class r_1 has just been decreased to zero level, then Lemma 4.3 implies that we can have two new controls that both keep classes $R \cup \{r_1\}$ customers at zero level. The first control uses full capacity and only works on $R \cup \{r_1\} \cup \{r_2\}$ classes of customers while the second control works under capacity and only works on $R \cup \{r_1\}$ classes of customers.

4.2.3 An Optimal Control Policy

In this section, we propose a control policy based on a sequence of linear programs. The solutions to these linear programs are can be used to define an optimal solution for the dual problem $(MBLP^*)$.

Consider the following control algorithm:

<u>Algorithm $\mathcal{B}$ $(B, a, b, w, x(0), N)$.</u>

1. $N^0 = N$. Solve the optimization problem

$$\begin{aligned} \min \quad & \eta^0 \\ \text{s. t.} \quad & B'\pi^0 + a\eta^0 \geq 0 \\ & \pi^0 = w \end{aligned}$$

<u>If</u> $\eta^0 < 0$, let $u(t) = 0$, and $\eta^0 = 0$ and stop the algorithm.

<u>Otherwise</u> let $s^0 = \arg\min_{i \in N^0}[B'\pi^0 + a\eta^0]_i$ and let u^0 be

$$u_i^0 = \begin{cases} 0, & \text{if } i \neq s^0 \\ \frac{1}{a_i}, & \text{otherwise.} \end{cases}$$

<u>If</u> s^0 never becomes zero, let $u(t) = u^0$ for all t and stop the algorithm.

<u>Otherwise</u> let t_0 be the first time in $[0, T]$ such that the queue length of class s^0 customers becomes zero, let $u(t) = u^0$ for all $0 \leq t \leq t_0$ and set

$$R^0 = \{s^0\}$$

$$N^0 = N \setminus \{s^0\}.$$

2. For $k = 1, 2, \cdots, n-1$ do

Solve the optimization problem

$$\begin{aligned} \min \quad & \eta^k \\ \text{such that} \quad & [B'\pi^k + a\eta^k]_{R^{k-1}} = 0 && (4.26) \\ & [B'\pi^k + a\eta^k]_{N^{k-1}} \geq 0 && (4.27) \\ & \pi^k_{N^{k-1}} = w_{N^{k-1}} && (4.28) \\ & \pi^k_{R^{k-1}} \leq \pi^{k-1}_{R^{k-1}} && (4.29) \\ & \eta^k \geq 0. && (4.30) \end{aligned}$$

If none of the inequalities in (4.27) becomes equality in the optimal solution, let R be $R^{k-1} \setminus \{s^{k-1}\}$, r_1 be s^{k-1} and r_2 be any element in N such that $r_2 \neq r_1$ and $r_2 \notin R$, by Lemma 4.3, there exists u^k (cf. $\bar{u}$) such that

$$\begin{aligned} [Bu^k + b]_{R^{k-1}} &= 0 \\ a'u^k &\leq 1 \\ u^k_{R^{k-1}} \geq 0, \quad u^k_{N^{k-1}} &= 0. \end{aligned}$$

We let $u(t) = u^k$ for all $t \geq t_{k-1}$, i.e., we do not work on any additional class, maintain the classes in R^{k-1} at zero level. Stop the algorithm.

Otherwise

Let $s^k = \arg\min_{i \in N^{k-1}} [B'\pi^k + a\eta^k]_i$. Let R be $R^{k-1} \setminus \{s^{k-1}\}$, r_1 be s^{k-1} and r_2 be s^k, by Lemma 4.3, there exists u^k (cf. $\tilde{u}$) such that

$$\begin{aligned} [Bu^k + b]_{R^{k-1}} &= 0 \\ a'u^k &= 1 \\ u^k_{R^{k-1} \cup \{r^2\}} \geq 0, \quad u^k_{N^{k-1} \setminus \{r^2\}} &= 0, \end{aligned}$$

i.e., we distribute the whole effort in such a way that customers in classes R^{k-1} are maintained at zero level and we only work on classes in $R^{k-1} \cup \{s^k\}$.

If s^k never becomes zero, let $u(t) = u^k$ for all $t \geq t_{k-1}$ and stop the algorithm.

Otherwise let t_k be the first time in $[0, T]$ such that the queue length of class s^k customers becomes zero, let $u(t) = u^k$ for all $t_{k-1} \leq t \leq t_k$ and set

$$R^k = R^{k-1} \cup \{s^k\}$$

$$N^k = N^{k-1} \setminus \{s^k\}.$$

There are two stages in Algorithm $\mathcal{B}$. At both stages of Algorithm B, whenever we have kept some classes of customers at zero level, we will pick another class of customers with the highest index $-\frac{[B'\pi^k]_i}{a_i}$ (we break ties arbitrarily) to serve depending on whether it is locally profitable to do so. The old classes of customers will be kept at the zero level and the newly picked customers receive all the remaining effort.

Theorem 4.1 *The proposed policy is optimal for $(MBLP)$. If the algorithm exits with no inequalities in* (4.27) *at equality in the optimal solution during the second stage of the algorithm, then the objective value of linear program is zero. For the case $w \geq 0$, the algorithm will not exit with $\eta^0 < 0$ and the constraint* (4.30) *is redundant.*

Proof We first prove that if there are no inequalities in (4.27) at equality in the optimal solution during the second stage of the algorithm, then $\eta^k = 0$. Suppose $\eta^k > 0$. By Lemma 4.3, there exists $\bar{x}$, such that

$$[B'\bar{x}]_{R^{k-1}} = 0, \qquad \bar{x}_{N^{k-1}} = w_{N^{k-1}}.$$

So

$$[B'x]_{R^{k-1}} = a_{R^{k-1}}, \qquad x_{N^{k-1}} = 0$$

is feasible (for example, $\frac{\bar{x}-\pi^{k-1}}{\eta^{k-1}}$ is a feasible solution). By Lemma 4.2, there exists ξ such that

$$[B'\xi]_{R^{k-1}} = a_{R^{k-1}}, \qquad \xi_{R^{k-1}} \leq 0, \qquad \xi_{N^{k-1}} = 0.$$

So, a small perturbation to π^k along the direction ξ (adjusting η^k accordingly) can be both feasible and lead to a strict decrease in the objective value. This is a contradiction.

We prove the optimality of the proposed policy by exhibiting a dual feasible solution, which together with the primal solution satisfies the complementary slackness conditions (4.7)–(4.9).

Let K be the largest k such that s^k is well defined in the proposed algorithm. We see $t_k \leq T$ is the first time in $[0,\ T]$ the queue length of class s^k customers becomes zero (if no such time exists, $t_k = T$). By Lemmas 4.2 and 4.4, both optimization problems in stage 1 and stage 2 are feasible and have finite optimal solutions and the choice of u^k therein is also valid. The primal variables can be specified according to Algorithm $\mathcal{B}$. Let $t_{-1} = 0$, for any $t \in [t_{i-1},\ t_i)$, $i = 0,\ 1,\ \cdots,\ K$, we define the dual variables in the following way:

$$\begin{aligned} \pi(t) &= \pi^i \\ \eta(t) &= \sum_{j=i+1}^{K} (t_j - t_{j-1})\eta^j + (t - t_{i-1})\eta^i. \end{aligned}$$

For $t \in [t_K,\ T]$, we let $\pi(t) = \eta(t) = 0$. It is easy to check that $(\pi(t), \eta(t))$ is feasible for the dual problem and for all $t \in [0,\ T]$, relations (4.7)–(4.9) hold.

Thus we have exhibited a dual feasible solution $(\pi(t), \eta(t))$ that together with the primal solution satisfies the complementary slackness condition. By Corollary 4.1, the primal solution is optimal.

Let j be the index such that w_j is maximized, from the j-th inequality of the linear program in the first stage, we see $\eta^0 \geq 0$. Consider the linear program in the second stage. Let j be the index such that π_j^k is maximized. $\pi_j^k \geq 0$ since N^{k-1} is not empty until s^{n-1} is defined. By the j-th relation among (4.26) and (4.27), the constraint $\eta^k \geq 0$ is redundant. □

4.2.4 Ties with Queueing Control Problem

Algorithm $\mathcal{B}$ does not depend on T. The optimization in both stages does not depend on b. If we are only interested in calculating the optimal indices, we can set $b = 0$, so the algorithm will not terminate solely due to not having enough work capacity. The calculation of the indices does not depends on the total capacity of the machine either. The order of the sequence $\{s^k\}$ will be the same if we scale a_i by the same factor. We remark that the indices produced by Algorithm $\mathcal{B}$ are closely connected to the dual algorithm in Bertsimas and Nino-Mora [15]. As shown in the following theorem, Algorithm $\mathcal{B}$ produces the same indices and can be considered as a generalization. Although Algorithm $\mathcal{B}$ applies to the deterministic problem, it can be applied to any finite horizon problem and the system need not have a steady state solution. Furthermore, no work conserving conditions are imposed. Before stating the theorem, we first describe the algorithm in Bertsimas and Nino-Mora [15].

Algorithm $\mathcal{C}$ (B, a, w, N).

Assume B is invertible, $a > 0$ and $w \geq 0$.

1. $\bar{N}^0 = N$,

$$q_i^{\bar{N}^0} = -{B'_{\bar{N}^0\bar{N}^0}}^{-1} a_{\bar{N}^0}$$

$$\begin{aligned} y^{\bar{N}^0} &= \min_{i \in \bar{N}^0} \frac{w_i}{q_i^{\bar{N}^0}} \\ \tau^n &= \arg\min_{i \in \bar{N}^0} \frac{w_i}{q_i^{\bar{N}^0}} \end{aligned}$$

2. For $k = 1,\ 2,\ \cdots,\ n-1$ do

$$\bar{N}^k = N \setminus \{\tau^n\}$$

$$q^{\bar{N}^k} = -{B'_{\bar{N}^k\bar{N}^k}}^{-1} a_{\bar{N}^k}$$

$$\begin{aligned} y^{\bar{N}^k} &= \min_{i \in \bar{N}^k} \frac{w_i - \sum_{j=0}^{k-1} q_i^{\bar{N}^j} y^{\bar{N}^j}}{q_i^{\bar{N}^k}} \\ \tau^{n-k} &= \arg\min_{i \in \bar{N}^k} \frac{w_i - \sum_{j=0}^{k-1} q_i^{\bar{N}^j} y^{\bar{N}^j}}{q_i^{\bar{N}^k}} \end{aligned}$$

3. Serve the customers according to the order $\tau^1, \tau^2, \cdots, \tau^n$.

Theorem 4.2 *If matrix B is invertible, $a > 0$ and $w \geq 0$, then Algorithms $\mathcal{B}$ and $\mathcal{C}$ produce the same optimal policy.*

Proof Since B is invertible and $w \geq 0$, by Lemma 4.2 and Theorem 4.1, constraints (4.29) and (4.30) are redundant. By Lemma 4.2, π^k and η^k are always nonnegative. η^k will be strictly positive unless $w = 0$ and so all the constraints in (4.27) are binding. So Algorithm $\mathcal{B}$ (the index calculation part) can be carried out without exiting prematurely from the loop. At every iteration in the second stage, there is a unique π^k associated with every η^k so that η^k and π^k are feasible for the linear program. So there is a unique optimal solution.

Given the output $\{\eta^k\}$ and $\{s^k\}$ of Algorithm $\mathcal{B}$, we will show that the sequence $\{s^k\}$ is a valid output of Algorithm $\mathcal{C}$ and $y^{\bar{N}^k} = \eta^{n-k-1} - \eta^{n-k}$ (We let $\eta^n = 0$ and $\pi^n = 0$ for notational convenience). In other words, $\{\eta^k\}$ is exactly the Generalized Gittin's index defined in Bertsimas and Nino-Mora [15]. We prove this by induction on the iteration of Algorithm $\mathcal{C}$.

At the first iteration of Algorithm $\mathcal{C}$, we consider the n-th iteration of Algorithm $\mathcal{B}$. We have

$$\begin{aligned} B'\pi^{n-1} + a\eta^{n-1} &= 0 \\ \pi^{n-1}_{s^{n-1}} &= w_{s^{n-1}} \\ \pi^{n-1} &\leq w. \end{aligned}$$

So

$$w \geq \pi^{n-1} = -\eta^{n-1} {B'_{\bar{N}^0\bar{N}^0}}^{-1} a = \eta^{n-1} q^{\bar{N}^0}.$$

Obviously, $\tau^n = s^{n-1} = \arg\min_{i \in \bar{N}^0} \frac{w_i}{q_i^{\bar{N}^0}}$ and $\eta^{n-1} = y^{\bar{N}^0}$. Now, assume that $s^{n-1}, \cdots, s^{n-k}$ coincide with $\tau^n, \cdots, \tau^{n-k+1}$ and $y^{\bar{N}^i} = \eta^{n-i-1} - \eta^{n-i}$ for $i = 0, \cdots, k-1$. At the $k+1$st iteration of Algorithm $\mathcal{C}$, we consider the $n-k$-th iteration of Algorithm $\mathcal{B}$,

$$\begin{aligned} [B'\pi^{n-k-1} + a\eta^{n-k-1}]_{\bar{N}^k} &= 0 \\ \pi^{n-k-1}_{\bar{N}^i \cup \{s^{n-k-1}\}} &= w_{\bar{N}^i \cup \{s^{n-k-1}\}} \\ \pi^{n-k-1} &\leq w. \end{aligned}$$

By the induction hypothesis, we have

$$\begin{aligned} [B'\pi^{n-i-1} + a\eta^{n-i-1}]_{\bar{N}^i} &= 0 \\ \pi^{n-i-1}_{\bar{N}^i \cup \{s^{n-i-1}\}} &= w_{\bar{N}^i \cup \{s^{n-i-1}\}} \\ \pi^{n-i-1} \leq w \qquad &\text{for } i < k. \end{aligned}$$

So

$$B'_{\bar{N}^k\bar{N}^k}(\pi^{n-k-1}_{\bar{N}^k} - \pi^{n-k}_{\bar{N}^k}) = -a_{\bar{N}^k}(\eta^{n-k-1} - \eta^{n-k}). \tag{4.31}$$

Similarly we have

$$B'_{\bar{N}^i\bar{N}^i}(\pi^{n-i-1}_{\bar{N}^i} - \pi^{n-i}_{\bar{N}^i}) = -a_{\bar{N}^i}(\eta^{n-i-1} - \eta^{n-i}) = -a_{\bar{N}^i}y^{\bar{N}^i} \qquad \text{for } i < k, \tag{4.32}$$

where the last equality is because of induction hypothesis. By (4.31) and (4.32), we have

$$w_{\bar{N}^k} \geq \pi^{n-k}_{\bar{N}^k} = (\eta^{n-k-1} - \eta^{n-k})q^{\bar{N}^k}_{\bar{N}^k} + \sum_{i=0}^{k-1} q^{\bar{N}^i}_{\bar{N}^i}y^{\bar{N}^i}$$

and $\pi^{n-k}_{s^{n-k-1}} = w_{s^{n-k-1}}$. So

$$\begin{aligned} \eta^{n-k-1} - \eta^{n-k} &= \min_{i\in\bar{N}^k} \frac{w_i - \sum_{j=0}^{k-1} q_i^{\bar{N}^j} y^{\bar{N}^j}}{q_i^{\bar{N}^k}} \\ s^{n-k-1} &= \arg\min_{i\in\bar{N}^k} \frac{w_i - \sum_{j=0}^{k-1} q_i^{\bar{N}^j} y^{\bar{N}^j}}{q_i^{\bar{N}^k}}. \end{aligned}$$

So $y^{\bar{N}^k} = \eta^{n-k-1} - \eta^{n-k}$ and we can pick $\tau^{n-k} = s^{n-k-1}$. This completes the induction. □

None of the conditions $a > 0$ and $b \geq 0$ can be removed. We make our decisions based on the information of the dual problem. This theorem shows, if we always use the locally optimal control law for the fluid problem, global optimality can be achieved for both the fluid problem and the corresponding stochastic problem. This shows great promise of the fluid model.

4.3 A Single Multiclass Queue with Separable Convex Quadratic Costs

In this section, we consider the optimal control for the following fluid flow approximation to a make-to-stock multiclass queue scheduling problem. The facilities produce according to customer demand, and completed jobs enter a finished goods inventory, which, in turn, services actual customer demand. The arrival rate of class i jobs is b_i

and the service rate of class i jobs is $\frac{1}{a_i}$. The goal is to dynamically decide which job class, if any, to serve next in order to minimize the average cost incurred per unit of time over the interval $[0,\ T]$, which includes separable convex quadratic backordering costs and holding costs for finished good inventory.

More specifically, we consider the problem

$$\begin{aligned} (MBQP) \quad & \textit{minimize} && \int_0^T \sum_{i=1}^{n} C_i(x_i(t))\, dt \\ & \textit{subject to} && x(t) = x(0) + \int_0^t (u(t) - b)\, dt \\ & && a'u(t) \leq 1 \\ & && u(t) \geq 0 \end{aligned}$$

where $C_i(x_i(t)) = w_i x_i(t)^2$, $w_i > 0$, for $i = 1, \cdots, n$. We assume $a > 0$, $b \geq 0$ and $a'b \leq 1$. $x(0)$ and $T \geq 0$ are given and fixed. $(MBQP)$ is not a linear fluid model, since the cost function is quadratic in the state variables which demonstrates that fluid models are quite powerful.

Unlike the linear cost case, we are not going to use duality theory for continuous linear programs to solve the problem. Instead, we first propose an index type policy, and then use a Lagrangian function to prove the optimality of the policy. We remark, however, that the essence of the approach (duality theory) is indeed the same.

Consider the following policy.

Policy $\mathcal{P}\ (a, b, w, x(0))$.

At any given time t,

Let

$$I^*(t) = \{\ i \mid i = \arg\min(\frac{2x_i(t)w_i}{a_i})\ \} \quad \text{and} \quad \alpha(t) = \sum_{i \in I^*(t)} \frac{a_i^2}{w_i}.$$

If $x_i(t) > 0$ for some $i \in I^*(t)$, let $u_i(t) = 0$ for all i.

Otherwise if $x_i(t) = 0$ for some $i \in I^*(t)$, let

$$u_i(t) = \begin{cases} b_i, & \text{for all } i \in I^*(t), \\ 0, & \text{otherwise.} \end{cases}$$

Otherwise let

$$u_i(t) = \begin{cases} b_i + \frac{1-\sum_{i\in I^*(t)} a_i b_i}{w_i} \cdot \frac{a_i}{\alpha(t)}, & \text{for all } i \in I^*(t), \\ 0, & \text{otherwise.} \end{cases} \quad (4.33)$$

In other words, we never work on any class that has positive inventory. Among the nonpositive inventory classes, we only work on those classes that have the smallest index $\frac{2x_i(t)w_i}{a_i}$. If there are no negative inventory classes, we work only to keep the zero inventory level classes at the zero inventory level. If there are negative inventory classes, we distribute the effort in the way specified by (4.33). Notice in (4.33), $\sum a_i u_i(t) = 1$ and for $i \in I^*(t)$, $u_i(t) - b_i$ is proportional to $\frac{a_i}{w_i}$. The intuition behind this is to keep $I^*(t)$ as stable as possible, if $i \in I^*(t)$ and there exist jobs that have backlogging, then $i \in I^*(\bar{t})$ for all $\bar{t} > t$.

Let $(u^*(t), x^*(t))$ be the feasible solution given by Policy $\mathcal{P}$. We define

$$\phi(t) = \min_{i\in[1,\cdots,n]} \{0, \frac{2x_i^*(t)w_i}{a_i}\}$$

and

$$V(t) = -\int_t^T \phi(t)\, dt.$$

From the definition, it is obvious that

$$\begin{aligned} V(t) &\geq 0 \quad \forall t \in [0,T] \\ V(T) &= 0 \end{aligned}$$

It is a fact that if $\phi(t) = 0$ for some t, the $\phi(\bar{t}) = 0$ for all $\bar{t} \geq t$. Consider the Lagrangian problem (OLP):

$$\begin{aligned} (OLP) \quad & \textit{minimize} \quad \int_0^T \sum_{i=1}^n C_i(x_i(t))\, dt + \int_0^T V(t)(a'u(t) - 1)\, dt \\ & \textit{subject to} \quad x(t) = x(0) + \int_0^t (u(t) - b)\, dt \\ & \qquad\qquad\quad u(t) \geq 0 \end{aligned}$$

Here, we used $V(t)$ to dualize the control constraint and the variable $V(t)$ serves as a Lagrange multiplier for the $(MBQP)$. (OLP) is equivalent to the following n

separate subproblems, for $i = 1, \cdots, n$:

$$
\begin{aligned}
(OLP_i) \quad & \text{minimize} && \int_0^T C_i(x_i(t))\,dt + \int_0^T a_i(u_i(t)-b_i)V(t)\,dt \\
& \text{subject to} && x_i(t) = x_i(0) + \int_0^t (u_i(t)-b_i)\,dt \\
& && u(t) \geq 0
\end{aligned}
$$

Lemma 4.1 *The solution $(u_i^*(t),\ x_i^*(t))$ given by Policy $\mathcal{P}$ is an optimal solution for (OLP_i).*

Proof Through integration by parts, we have

$$
\begin{aligned}
& \int_0^T C_i(x_i(t))\,dt + \int_0^T a_i(u_i(t)-b_i)V(t)\,dt \\
= & \int_0^T C_i(x_i(t))\,dt + \int_0^T a_i V(t)\,dx_i(t) \\
= & \int_0^T C_i(x_i(t))\,dt + a_i V(t)x_i(t)|_0^T - \int_0^T a_i x_i(t)\dot{V}(t)\,dt \\
= & \int_0^T C_i(x_i(t))\,dt - a_i V(0)x_i(0) - \int_0^T a_i x_i(t)\dot{V}(t)\,dt
\end{aligned}
$$

So (OLP_i) is the same as:

$$
\begin{aligned}
& \text{minimize} && \int_0^T C_i(x_i(t))\,dt - \int_0^T a_i x_i(t)\dot{V}(t)\,dt \\
& \text{subject to} && x_i(t) = x_i(0) + \int_0^t (u_i(t)-b_i)\,dt \\
& && u(t) \geq 0,
\end{aligned}
$$

which leads by adding and subtracting $\frac{a_i^2}{4w_i^2}\dot{V}(t)^2$ to:

$$
\begin{aligned}
& \text{minimize} && \int_0^T [w_i(x_i(t) - \frac{a_i}{2w_i}\dot{V}(t))^2 + \frac{a_i^2}{4w_i^2}\dot{V}(t)^2]\,dt \\
& \text{subject to} && x_i(t) = x_i(0) + \int_0^t (u_i(t)-b_i)\,dt \\
& && u(t) \geq 0,
\end{aligned}
$$

which is the same as

$$\begin{aligned}
&\textit{minimize} && \int_0^T w_i(x_i(t) - \frac{a_i}{2w_i}\dot{V}(t))^2 \, dt \\
&\textit{subject to} && x_i(t) = x_i(0) + \int_0^t (u_i(t) - b_i) \, dt \\
& && u(t) \geq 0.
\end{aligned}$$

We will show that solution $(u_i^*(t),\ x_i^*(t))$ given by Policy $\mathcal{P}$ satisfies for all $t \in [0,\ T]$

$$x_i^*(t) = \arg \min_{x_i(t) \text{ feasible}} (x_i(t) - \frac{a_i}{2w_i}\dot{V}(t))^2 \tag{4.34}$$

This is because if $\dot{V}(0) = 0$, then $\dot{V}(t) = 0$ for all $t > 0$. Policy $\mathcal{P}$ lets class i jobs decrease at full speed b_i before it hits zero and maintain class i jobs at the zero level afterwards, so (4.34) is satisfied. If $\dot{V}(t) < 0$, let t_i be the smallest time t such that $i \in I^*(t)$ (if no such t exists, let $t_i = T$). For $t \leq t_i$, Policy $\mathcal{P}$ lets class i jobs decrease at full speed b_i, so $x_i(t) > x_i^*(t) \geq \frac{a_i}{2w_i}\dot{V}(t)$ for any feasible $x_i(t)$ and (4.34) is satisfied for $t \leq t_i$. For $t > t_i$, $i \in I^*(t)$ and so $x_i^*(t) = \frac{a_i}{2w_i}\dot{V}(t)$ and (4.34) is trivially satisfied. Since $(u_i^*(t),\ x_i^*(t))$ is also feasible for the last non-linear program. So it solves it. □

Theorem 4.3 *The proposed policy is optimal for* $(MBQP)$.

Proof We prove this by showing the multiplier $V(t)$ together with the primal solution satisfies the complementary slackness condition $V(t)(a'u(t) - 1) = 0$ for all t. It is easily seen that the optimal solution value for problem $(MBQP)$ is always greater than that of (OLP). By Lemma 4.1, the primal solution $(u^*(t),\ x^*(t))$ solves the problem (OLP). So if we can show $V(t)(a'u^*(t) - 1) = 0$ for all t, this primal solution $(u^*(t),\ x^*(t))$ actually attains the minimum solution value of $(MBQP)$ and thus is optimal for $(MBQP)$. We now verify the complementary slackness condition. By definition, $V(t) \geq 0$. If $V(t) = 0$, the condition obviously holds. $V(t) > 0$ only when $\phi(t) < 0$. In this case (4.33) holds which implies $a'u^*(t) = 1$. □

The policy distributes the service effort among the classes that have the best index (the smallest nonpositive $\frac{2x_i^*(t)w_i}{a_i}$). The indices depend on the states and thus are dynamic.

4.4 Single-Class Queueing Network, the Tandem Case

In this section, we analyze the optimal solution structure for tandem queues and its close extensions. These problems have important applications in manufacturing, such as production of automobiles (see Buzacott and Shanthikumar [18]). Perkins and Kumar [71] considered pull manufacturing systems, where the objective is to minimize the sum of buffer holding costs and system shortfall/inventory costs, subject to an exogenous demand of constant rate. The fluid flow shop they considered is a special case of one of the problems we consider in this section.

4.4.1 Tandem Queues with Simple Feedback

The first problem we consider is the following. We have from left to right a series of workstations 1, $\cdots$, n in tandem as shown in Figure 4-2. Each workstation has its own unlimited buffer. Workstation i can produce at a rate up to $c_i > 0$ for $i = 1, \cdots, n$. The output of station i goes into the station $i+1$'s buffer and serves as the material for station $i+1$ for $i < n$. A fraction p of the output of station n goes to station 1 and $1-p$ exits the system. Initially, there are $x_i(0)$ material (or inventory) at station i. There are no external arrivals to the system. There is a cost of $w_i \geq 0$ per unit of time incurred for the inventory in workstation i. No negative inventory is allowed at any workstation. Our goal is to dynamically decide for each workstation, when to idle and how to work so as to minimize the average cost per unit of time over a time interval $[0,\ T]$. We can formulate our problem as the following continuous linear program.

$$
\begin{aligned}
(LCLP) \quad & \textit{minimize} && \int_0^T w'x(t)\ dt \\
& \textit{subject to} && x(t) = x(0) + \int_0^t Bu(t)\ dt \\
& && 0 \leq u(t) \leq c \\
& && x(t) \geq 0,
\end{aligned}
$$

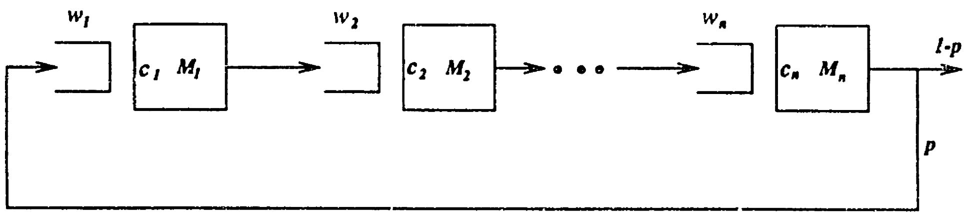

Figure 4-2: A series of queues in tandem with simple feedback

where $x(0) \geq 0$, $c > 0$, $w \geq 0$, $T \geq 0$ and

$$B = \begin{pmatrix} -1 & & & & p \\ 1 & -1 & & & \\ & 1 & \ddots & & \\ & & \ddots & -1 & \\ & & & 1 & -1 \end{pmatrix}.$$

We can easily transform the problem $(LCLP)$ into an equivalent separated continuous linear program. Just substitute $x(t)$ in the objective function by the right hand side of the first constraint and use integration by parts. A direct consequence of Propositions 2.1, 2.2 and 2.3 is

Corollary 4.2 *There always exists an optimal basic feasible solution to problem $(LCLP)$ whose control is piecewise constant. Let $(u(t), x(t))$ be a basic feasible solution to $(LCLP)$ whose control is piecewise constant, then for almost all $t \in [0, T]$, the following holds:*

1. *If $i < j$, $x_i(t) > 0$ and $x_j(t) > 0$ but $x_k(t) = 0$ for all $i < k < j$, then either $u_i(t) = 0$ or $u_i(t) = \min_{k \in [i, j-1]} c_k$.*

2. *If $i \geq j$, $x_i(t) > 0$ and $x_j(t) > 0$ but $x_k(t) = 0$ for all $i < k$ and $k < j$, then either $u_i(t) = 0$ or $u_i(t) = \min\{\min_{k \in [i,n]} c_k,\ \min_{k \in [1, j-1]} pc_k\}$.*

For the two stations i and j in the corollary, we call j the next positive inventory station with respect to station i. Obviously, a unique feasible solution $x(t)$ is determined by any feasible $u(t)$. By Proposition 2.3, there exists an optimal control $u(t)$ which is piecewise constant and has finite (albeit potentially exponential) number of breakpoints. Exploiting the particular structure of the problem, we will show there exists an optimal solution for $(LCLP)$ that has the following property:

Structural Property 1

There exist t_i, $i = 1, \cdots, n$, such that workstation i idles before t_i. At any time $t > t_i$ but before $x_i(t) = 0$, let j be the next positive inventory station with respect to station i, $u_i(t) = \min_{k \in [i,j-1]} c_k$ if $j > i$ and $u_i(t) = \min\{\min_{k\in[i,n]} c_k, \ \min_{k\in[1,j-1]} pc_k\}$ otherwise. After the inventory at workstation i hits zero, the incoming flow to the workstation will always equal to the out going flow, and the inventory level of the workstation will remain at zero.

In other words, workstation i starts working at time t_i. It works at the largest possible rate such that no other workstations that already have zero inventory level will become positive again. These n times $\{t_i\}$ specify a unique control policy $u(t)$ that has at most $2n$ constant control pieces, i.e., $u(t)$ will only change when either there is a workstation that starts working or there is a workstation whose inventory level becomes zero.

The dual of $(LCLP)$ can be written as

$$\begin{array}{lll} (LCLP^*) & \textit{maximize} & \int_0^T (x(0)'\pi(t) - c'\eta(t)) \, dt \\ & \textit{subject to} & \eta(t) + \int_t^T B'\pi(t) \, dt \geq 0 \\ & & \pi(t) \leq w \\ & & \eta(t) \geq 0. \end{array}$$

Arguing as in Chapter 3, weak duality holds between the primal and dual programs. Moreover, strong duality holds if a primal-dual optimal solution pair satisfies the following complementary slackness condition:

$$\begin{aligned} x(t)'(\pi(t) - w) &= 0 \\ (u(t) - c)'\eta(t) &= 0 \end{aligned}$$

$$u(t)' \left[\int_t^T B'\pi(t)\ dt + \eta(t) \right] = 0 \qquad \text{for all } t. \tag{4.35}$$

In order to prove Structural Property 1, we make for convenience the following assumptions. We will remove them later in the section (cf. Corollary 4.3).

Assumption 4.1

1. $0 < p < 1$
2. $w_i \neq w_j$ *if* $i \neq j$
3. $w_i \neq pw_j$
4. $x(0) > 0$.

First, we prove the following property:

Lemma 4.5 *Let* $(u(t), x(t))$ *be a basic optimal solution to* $(LCLP)$ *with* $u(t)$ *piecewise constant. Under Assumption* 4.1, *if* $x_i(t) > 0$ *and station* j *is the next positive inventory station with respect to station* i *for some* t, *then*

1. *If* $i < j$ *and* $w_i > \min_{k\in[i,j]} w_k$ *then* $u_i(t) > 0$.
2. *If* $i \geq j$ *and* $w_i > \min\{\min_{k\in[i,n]} w_k, \min_{k\in[1,j]} pw_k\}$ *then* $u_i(t) > 0$.

Proof We first prove the first part of the lemma. Suppose the contrary. By assumption, $i > j$, $w_i > \min_{k\in[i,j]} w_k$, there exists $0 < t_1 < t_2 < T$ such that $x_i(t_1) > 0$ but $u_i(t) = 0$ for $t \in [t_1,\ t_2]$. Let $\bar{j} = \arg\min_{k\in[i,j]} w_k$. Consider the following feasible solution $(\tilde{u}(t),\ \tilde{x}(t))$ for $(LCLP)$. $\tilde{u}_k(t)$ equals to $u_k(t)$ for all $k \notin [i, \bar{j}-1]$. We let

$$\tilde{x}_i(t) = \begin{cases} x_i(t), & \text{for } t \leq t_1; \\ x_i(t_1) - (t - t_1)\delta, & \text{for } t \in (t_1, t_2); \\ \max\{x_i(t) - \delta, 0\}, & \text{for } t \geq t_2, \end{cases}$$

where $\delta < \min\{x_i(t_1), (t_2 - t_1) \min_{k\in[i,\bar{j}-1]} c_k\}$. We see that $\tilde{x}_i(t)$ uniquely specifies a control $0 \leq \tilde{u}_i(t)$ and $\tilde{u}_i(t) \leq u_i(t)$ for $t \geq t_2$. For $k \in [i+1, \bar{j}-1]$, we recursively let

$$\tilde{u}_k(t) = \begin{cases} u_k(t), & \text{for } t \leq t_1; \\ \frac{\delta}{t_2 - t_1}, & \text{if } t \in (t_1, t_2); \\ u_k(t), & \text{if } t \geq t_2 \text{ and } \tilde{x}_k(t) > 0; \\ \max\{0, \tilde{u}_{k-1}(t) + u_k(t) - u_{k-1}(t)\}, & \text{if } t \geq t_2 \text{ and } \tilde{x}_k(t) = 0. \end{cases}$$

We see that $0 \leq \tilde{u}_k(t) \leq u_k(t)$ for all $t \geq t_2$. It can be shown by induction on k that $\tilde{x}_k(t) \leq x_k(t)$ for all $k \in [i, \bar{j}-1]$ and for all t. Since $\sum_{k \in [i,\ \bar{j}]} x_k(t) = \sum_{k \in [i,\ \bar{j}]} \tilde{x}_k(t)$, it is easy to see that $(\tilde{u}(t), \tilde{x}(t))$ has strictly smaller cost than $(u(t), x(t))$ and we have a contradiction. This proves the first part of the lemma.

We now prove the second part of the lemma. If $i = j$, $x_i(t_1) > 0$ but $u_i(t) = 0$ for all $t \in (t_1, t_2)$, we have a contradiction, since the new policy that sends some flow δ from station i to station i through station n during (t_1, t_2) and uses $\frac{x_i(t_1) - (1-p)\delta}{x_i(t_1)} u(t - t_2 + t_1)$ after t_2 would have a smaller cost. So, we assume $i > j$. When there exists a station $\bar{j} > i$ such that $\bar{j} = \arg\min_{k \in [i,n]} w_k$, exactly the same proof as the one for the first part of the lemma applies. So we can assume $w_i = \arg\min_{k \in [i,n]} w_k$. Let $\bar{j} = \arg\min_{k \in [1,j]} w_k$. Suppose the second part of the lemma is not true, by assumption $w_i > p w_{\bar{j}}$, there exists $0 < t_1 < t_2 < T$ such that $x_i(t_1) > 0$ but $u_i(t) = 0$ for $t \in [t_1,\ t_2]$. Consider the following feasible solution $(\tilde{u}(t),\ \tilde{x}(t))$ for $(LCLP)$. $\tilde{u}_k(t)$ equals to $u_k(t)$ for all $k \in [\bar{j}, i-1]$. We let

$$\tilde{x}_i(t) = \begin{cases} x_i(t), & \text{for } t \leq t_1; \\ x_i(t_1) - (t - t_1)\delta, & \text{for } t \in (t_1, t_2); \\ \max\{x_i(t) - \delta, 0\}, & \text{for } t \geq t_2, \end{cases}$$

where $\delta < \min\{x_i(t_1), (t_2 - t_1) \min_{k \in [i, \bar{j}-1]} c_k\}$. We see that $\tilde{x}_i(t)$ uniquely specifies a control $0 \leq \tilde{u}_i(t)$ and $\tilde{u}_i(t) \leq u_i(t)$ for $t \geq t_2$. For $k \in [i+1, n]$, we recursively let

$$\tilde{u}_k(t) = \begin{cases} u_k(t), & \text{for } t \leq t_1; \\ \frac{\delta}{t_2 - t_1}, & \text{if } t \in (t_1, t_2); \\ u_k(t), & \text{if } t \geq t_2 \text{ and } \tilde{x}_k(t) > 0; \\ \max\{0, \tilde{u}_{k-1}(t) + u_k(t) - u_{k-1}(t)\}, & \text{if } t \geq t_2 \text{ and } \tilde{x}_k(t) = 0. \end{cases}$$

We let

$$\tilde{u}_1(t) = \begin{cases} u_1(t), & \text{for } t \le t_1; \\ \frac{p\delta}{t_2 - t_1}, & \text{if } t \in (t_1, t_2); \\ u_1(t), & \text{if } t \ge t_2 \text{ and } \tilde{x}_1(t) > 0; \\ \max\{0, p\tilde{u}_n(t) + u_1(t) - pu_n(t)\}, & \text{if } t \ge t_2 \text{ and } \tilde{x}_1(t) = 0, \end{cases}$$

and for $k \in [2, \bar{j} - 1]$, we recursively let

$$\tilde{u}_k(t) = \begin{cases} u_k(t), & \text{for } t \le t_1; \\ \frac{p\delta}{t_2 - t_1}, & \text{if } t \in (t_1, t_2); \\ u_k(t), & \text{if } t \ge t_2 \text{ and } \tilde{x}_k(t) > 0; \\ \max\{0, \tilde{u}_{k-1}(t) + u_k(t) - u_{k-1}(t)\}, & \text{if } t \ge t_2 \text{ and } \tilde{x}_k(t) = 0. \end{cases}$$

We see that $0 \le \tilde{u}_k(t) \le u_k(t)$ for all $t \ge t_2$. It can be shown by induction on k that $\tilde{x}_k(t) \le x_k(t)$ for all $k \notin [\bar{j}, i-1]$ and $\sum_{k \in [\bar{j}+1, i-1]} x_k(t) = \sum_{k \in [\bar{j}+1,\, i-1]} \tilde{x}_k(t)$ for all t. It is easy to see that $(\tilde{u}(t), \tilde{x}(t))$ has strictly smaller cost than $(u(t), x(t))$ and we have a contradiction. This proves the second part of the lemma. □

Lemma 4.6 *Let Assumption* 4.1 *holds. Let* $(u(t), x(t))$ *be an optimal solution to* $(LCLP)$ *with* $u(t)$ *piecewise constant. If for* $0 \le \tilde{t}_i < \bar{t}_i \le T$, $u_i(t) = 0$ *for* $t \in (\tilde{t}_i,\ \bar{t}_i)$ *and* $x_i(\tilde{t}_i) > 0$, *then* $u(t) = 0$ *for all* $t < \tilde{t}_i$.

Proof Suppose the contrary, let station j be the next positive inventory station with respect to station i at time $\tilde{t}_i$. By Lemma 4.5, we see $w_i = \min_{k \in [i,j]} w_k$ if $i < j$ and $w_i = \min\{p \min_{k \in [1,j]} w_k, \min_{k \in [i,n]} w_k\}$ otherwise. Since station i has been busy before $\tilde{t}_i$, there are some flows going from i to j (these flows may stay between $i+1$ and $j-1$ or between $[i+1, n] \cup [1, j-1]$ for some time) before $\tilde{t}_i$. Since there is only a single class of customers, we can always assume every workstation adopts a FCFS type of policy. Under this policy, it is easy to see that there exists $\delta > 0$ inventory in station j at epoch $\tilde{t}_i$ came from station i, otherwise station j would not have positive inventory at epoch $\tilde{t}_i$.

If $i < j$ consider a new policy that withholds δ unit (so it can be sent to station j within time interval $[\tilde{t}_i,\ \bar{t}_i]$) of flow at station i before $\tilde{t}_i$ and sends this amount of flow to station j during $[\tilde{t}_i,\ \bar{t}_i]$. The only change in the state variable is the flows among the stations from i to j. More precisely, this small amount of flow will change from staying among the stations from $i+1$ to j for some positive amount of time to spending this time at station i instead. Obviously, this new policy is strictly better than the original policy and this leads to a contradiction.

If $i \geq j$, consider a new policy that withholds $\frac{\delta}{p}$ unit (so it can be sent to station j through station n within time interval $[\tilde{t}_i,\ \bar{t}_i]$) of flow at station i before $\tilde{t}_i$ and send this amount of flow to station j via station n during $[\tilde{t}_i,\ \bar{t}_i]$. The only change in the state variable is the flows among the stations from i to n and from 1 to j. More precisely, this $\frac{\delta}{p}$ unit of flow will change from staying among the stations in $[i+1,n] \cup [1,j]$ (and possibly staying outside the systems) for some positive amount of time to spending this time at station i instead. This new policy is strictly better than the original policy and this also leads to a contradiction. □

Let i and j be two workstations. We say station j is a bottleneck with respect to station i if there exists a station l, such that one of the following three conditions holds:

$$
\begin{aligned}
w_i &= \min_{k\in[i,\ l-1]} w_k, \qquad c_i > \min_{k\in[j,\ l-1]} c_k, \qquad i<j<l \\
w_i &= \min\{\min_{k\in[i,\ n]} w_k, p\min_{k\in[1,\ l-1]} w_k\}, \quad c_i > \min\{\min_{k\in[j,\ n]} c_k, \frac{1}{p}\min_{k\in[1,\ l-1]} c_k\}, \quad l \leq i < j \\
w_i &= \min\{\min_{k\in[i,\ n]} w_k, p\min_{k\in[1,\ l-1]} w_k\}, \quad c_i > \frac{1}{p}\min_{k\in[j,\ l-1]} c_k, \quad j \leq l \leq i
\end{aligned} \tag{4.36}
$$

Lemma 4.7 *Suppose Assumption* 4.1 *holds. If station j is a bottleneck with respect to station i, then in an optimal solution $(u(t), x(t))$ to the problem $(LCLP)$, we have*

$$
\begin{aligned}
u_i(t) &\leq \min_{k\in[j,\ l-1]} c_k, \quad \text{if } \ i<j<l \\
u_i(t) &\leq \min\{\min_{k\in[j,\ n]} c_k, \frac{1}{p}\min_{k\in[1,\ l-1]} c_k\}, \quad \text{if } \ l\leq i<j \\
u_i(t) &\leq \frac{1}{p}\min_{k\in[j,\ l-1]} c_k, \quad \text{if } \ j\leq l\leq i
\end{aligned}
$$

where station l is a station that satisfies one of the conditions in (4.36).

Proof We only give the proof for the case $i<j<l$, the other two cases can be proven similarly. Suppose station i works at rate $\bar{c}_i > \min_{k\in[j,\ l-1]} c_k$ during $[t_1,\ t_2]$, where $t_1 < t_2$. Let $\bar{k} = \arg\min_{k\in[i,\ l-1]} c_k$. Obviously $i < \bar{k}$. Denote $u_{\bar{k}}(t)$ as the control for station $\bar{k}$ in the old policy. We know $u_{\bar{k}}(t)$ is no greater than $c_{\bar{k}}$. Same as in the proof of Lemma 4.6, we assume each workstation adopts a FCFS type of policy. Denote t_3 as the first time in the policy station $\bar{k}$ output material sent from station i after t_1. Consider a new control policy such that workstation i idles during $[t_1,\ t_3]$ and works at rate $u_{\bar{k}}(t)$ starting from t_3. We let the control of all the other workstations be the same as the old control, except for the workstations in $[i+1,\ \bar{k}-1]$. During $[t_1,\ t_3]$, these workstations use the same control as the old control policy whenever possible. After t_3, we let all these workstations working at rate $u_{\bar{k}}(t)$. By FCFS discipline, we know workstations in $[i+1,\ \bar{k}]$ are all occupied at t_3 by materials that are sent from station i after time t_1. Since workstation i idles during $[t_1,\ t_3]$ in the new control, the workstations in $[i+1,\ \bar{k}-1]$ are empty after t_3 in the new control policy. For the workstations in $[i+1,\ \bar{k}]$ during $[t_1,\ T]$, the changes in control can be viewed as taking all the extra material sent from workstation i after t_1 away from these workstations and put them back at station i. The inventory level at any other workstation will not change. The new policy is strictly better than the old control policy. This leads to a contradiction. □

We are now ready to prove the following key lemma, which will lead to the structural property we need.

Lemma 4.8 *Suppose Assumption* 4.1 *holds. There is an optimal solution, such that after the inventory of a workstation hits zero, it will remain at zero level.*

Proof By Corollary 4.2, there is an optimal basic feasible $(\bar{x}(t), \bar{u}(t))$ such that $\bar{u}(t)$ is piecewise constant. Suppose the lemma does not hold, let t_1 be a smallest time such that there are stations that become positive inventory again. Let station j be a station that has positive inventory at time t_1. This workstation certainly exists, otherwise there wont be any workstation that becomes positive inventory again at time t_1. Let station i be the first workstation in the downstream of j that becomes positive inventory again at time t_1. Let l be the nearest station in the downstream of station i that either becomes positive again at time t_1 or whose inventory level is positive at time t_1. There are three possibilities, namely $j < i < l$, $i < l \le j$ and $l \ge j > i$. Here we only give the proof for the case $j < i < l$, the other two cases are similar.

Let $t_1^- < t_1$ and $t_1^+ > t_1$ be two times such that $\bar{u}(t)$ is constant over $[t_1^-, t_1)$ and

$[t_1, t_1^+]$. Since station i becomes positive inventory again at t_1, by Corollary 4.2, we have $\bar{u}_{i-1}(t_1^+) = \min_{k\in[j,\ i-1]} c_k$. Since station i becomes positive inventory again at t_1, by Lemma 4.6, it must be busy immediately after time t_1. Corollary 4.2 again implies:

$$\bar{u}_i(t_1^+) = \min_{k\in[i,\ l-1]} c_k. \tag{4.37}$$

So, we have

$$\min_{k\in[j,\ i-1]} c_k > \min_{k\in[i,\ l-1]} c_k. \tag{4.38}$$

We consider three cases.

Case 1. There exists a station $m \in [j,\ i-1]$ such that $w_m = \min_{k\in[j,\ l]} w_k$. By (4.38), station i is a bottleneck with respect to station m. By Lemma 4.7,

$$\bar{u}_{i-1}(t) \leq \bar{u}_m(t) \leq \min_{k\in[m,l-1]} c_k$$

for all $t \in [t_1, t_1^+]$. In view of (4.37), we have a contradiction.

Case 2. $w_i = \min_{k\in[j,\ l]} w_k$. Since all the stations in $[j, i-1]$ will always send as much flow as possible to station i, we see the incoming flow to station i will not increase from time t_1^- to t_1^+. So the out going flow of station i has strictly decreased immediately after t_1, otherwise the inventory of station i will not become positive again. By (4.37), the service rate of station i in $[t_1^-, t_1]$ must be greater than $\min_{k\in[i,\ l-1]} c_k$. However, since the inventory of all the workstations in $[i+1, l-1]$ has either just dropped to zero at t_1 or has stayed at zero during $[t_1^-, t_1]$, we have

$$\bar{u}_i(t) \leq \min_{k\in[i,\ l-1]} c_k$$

for $t \in [t_1^-, t_1]$. This is a contradiction.

Case 3. There exists a station $m \in [i+1,\ l]$ such that $w_m = \min_{k\in[j,\ l]} w_k$. If there is a workstation $h \in [j,\ i]$ such that $w_h < w_j$. We see, the output of workstation j will be no less than $\min_{k\in[j,\ h-1]} c_k$ before t_1. So the incoming flow of station i at t_1^- will be no less than $\min_{k\in[j,\ h-1]} c_k$ which is in turn no less than $\min_{k\in[j,\ i-1]} c_k$. So the incoming flow of station i will not increase from time t_1^- to t_1^+. By the same argument as *Case 2*, we can exhibit a contradiction. So,

$$w_j = \min_{k\in[j,\ i]} w_k \tag{4.39}$$

Let station q be the first workstation in $[i+1,\ m]$ such that $w_q < w_j$. Now consider the segment $[j,\ q]$. We have from (4.39) that

$$w_j = \min_{k\in[j,\ q-1]} w_k. \tag{4.40}$$

It is easy to see that $\bar{u}(t_1^+) \geq \min_{k\in[i,\ q-1]} c_k$. In view of (4.37), we have

$$\min_{k\in[i,\ q-1]} c_k = \min_{k\in[i,\ l-1]} c_k.$$

By (4.38) we have

$$\min_{k\in[j,\ i-1]} c_k > \min_{k\in[i,\ q-1]} c_k. \tag{4.41}$$

(4.40) and (4.41) imply that station i is a bottleneck with respect to station j. By Lemma 4.7, we have $\bar{u}_j(t_1^+) \leq \min_{k\in[i,\ q-1]} c_k$. So the inventory of station i will not become positive again at time t_1 and we have a contradiction. □

By using Lemmas 4.6 and 4.8, we can now prove the following structural property for problem $(LCLP)$.

Theorem 4.4 *For the tandem queues with simple feedback described by* $(LCLP)$*, if Assumption* 4.1 *holds, there exists an optimal solution* $u(t)$ *that has at most* $2n$ *constant pieces and Structural Property 1 holds.*

Proof Consider a basic optimal solution $(u(t), x(t))$ to $(LCLP)$. Denote the first time workstation i starts working as t_i. Obviously, workstation n never works. By Lemma 4.6, workstation i will work whenever possible after t_i. By Lemma 4.8, once the inventory level at station i hits zero, it will stay at zero. After t_i but before the inventory of station i hits zero, let j be the next positive inventory station with respect to station i at time t, by Corollary 4.2, $u_i(t) = \min_{k\in[i,j-1]} c_k$ if $j > i$ and $u_i(t) = \min\{\min_{k\in[i,n]} c_k,\ \min_{k\in[1,j-1]} pc_k\}$ otherwise. After the inventory of station i hits zero, by Lemma 4.8, the working rate of workstation i is modulated by the incoming flow rate, i.e., the incoming flow to the workstation will always equal to the out going flow, so that the inventory level will remain at the zero level. By this argument, we see the working rate of a workstation changes only when there is either a workstation starts working or there is a workstation whose inventory level hits zero. Once $t_1, \cdots, t_n$ are given, a unique optimal basic feasible solution is determined. Thus the theorem is proved. □

Corollary 4.3 *Theorem* 4.4 *holds without Assumption* 4.1.

Proof Any problem that violates Assumption 4.1 can be approximated by a series of problems that satisfies these assumptions, due to the compactness of both the proposed optimal solution (i.e., the finiteness of the number of control pieces) and the feasible control set, we can prove the property for the original problem using limiting arguments. □

The second problem is a fluid tandem queueing control problem. It is a special case of the first problem with $p = 0$ and station n being a dummy station which never works and only serves as an unlimited buffer. The problem can be formulated as follows

$$\begin{aligned} (LCLP1) \quad \textit{minimize} \quad & \int_0^T w'x(t)\,dt \\ \textit{subject to} \quad & x(t) = x(0) + \int_0^t Bu(t)\,dt \\ & 0 \le u(t) \le c \\ & x(t) \ge 0, \end{aligned}$$

where $x(0) \ge 0$, $c > 0$, $w \ge 0$, $T \ge 0$ and B is the negative node-arc incidence matrix for the following line digraph: Node i of the graph corresponds to machine i and the edges are $(i,\ i+1)$ for $i = 1, \cdots, n-1$. The system is shown in Figure 4-3. The next theorem concerns the structural property of an optimal solution to this tandem queueing network.

Theorem 4.5 *For the fluid tandem queueing control problem described by* $(LCLP1)$, *Structural Property 1 holds.*

Example 4.1. Consider problem $(TAND)$ given in Section 1.1.1. Let $T = 6$, $\lambda = 0$, $\mu_1 = \frac{2}{5}$, $\mu_2 = \frac{6}{7}$, $x_1(0) = x_2(0) = 2$, $w_1 = 1$ and $w_2 = 2$. This is a problem $(LCLP)$

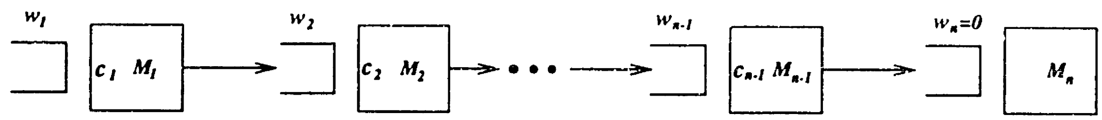

Figure 4-3: A series of queues in tandem

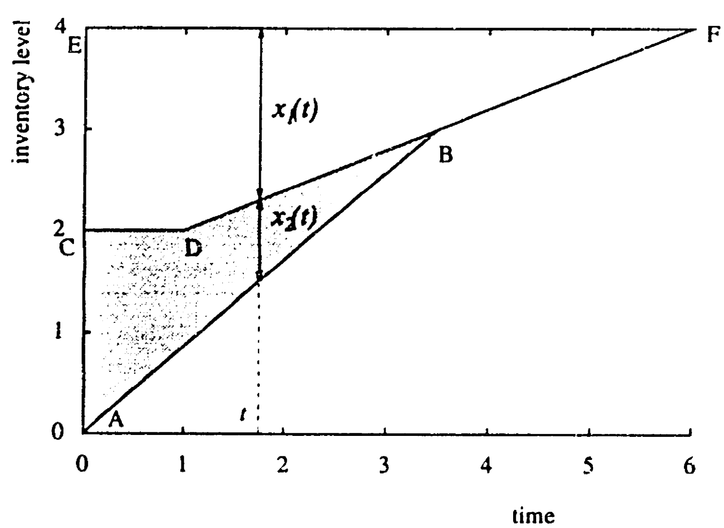

Figure 4-4: An optimal solution for Example 4.1

with $p = 0$. An optimal solution is shown in Figure 4-4. The station 1 idles during $[0, 1)$. It works at rate $\frac{2}{5}$ during $[1, 6]$. The station 2 works at rate $\frac{6}{7}$ during $[0, 3.5)$ and works at rate $\frac{2}{5}$ during $[3.5, 6]$. In Figure 4-4, the horizontal axis represents time and the vertical axis represents the inventory level for the two stations. The vertical distance between line segments $CDBF$ and EF represents the inventory level for station 1. The vertical distance between line segments ABF and $CDBF$ represents the inventory level for station 2.

Given the above results, a natural question would be: does Structural Property 1 for the tandem queueing network (i.e., ($LCLP$)) generalize to fluid approximation for Generalized Jackson networks? In the following, we give a four workstation Generalized Jackson network. In the optimal control of its fluid approximation, the inventory level at certain workstation can drop to zero, stay at zero for a while and then become positive again.

Example 4.2. Consider the network shown in Figure 4-5. Fraction 0.7 of the flow from

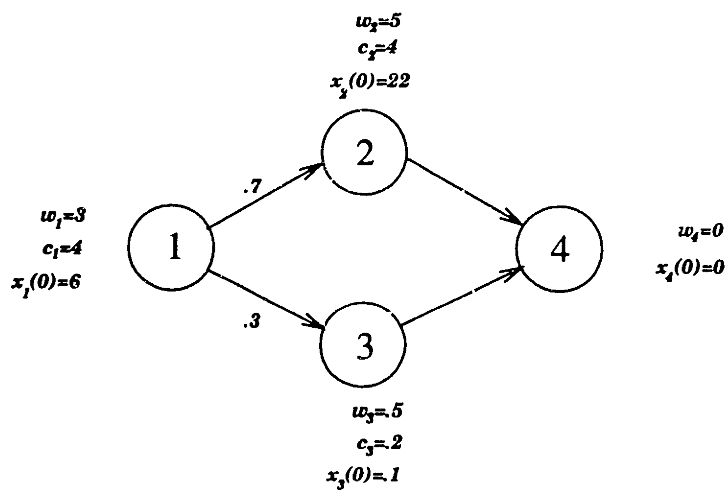

Figure 4-5: A four workstation queueing network

workstation 1 is sent to station 2 while 0.3 is sent to workstation 3. The flows can also be sent from workstations 2 and 3 to workstation 4. No other routes are allowed. The capacity for these workstations are: $c_1 = 4,\ c_2 = 4,\ c_3 = 0.2$. The holding cost per unit time at these workstations are: $w_1 = 3,\ w_2 = 5,\ w_3 = 0.5,\ w_4 = 0$. Initial inventories at these workstations are: $x_1(0) = 6,\ x_2(0) = 22,\ x_3(0) = 0.1,\ x_4(0) = 0$. $T = 39$. No backordering is allowed.

The transition matrix B in $(LCLP)$ for this problem is:

$$B = \begin{pmatrix} -1 & 0 & 0 \\ .7 & -1 & 0 \\ .3 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

A pair of optimal primal and dual solutions are given as follows. Denote

$$\begin{aligned} \mathcal{I}_1 &= [0,\ 0.5), \quad \mathcal{I}_2 = [0.5,\ 3\frac{349}{910}), \quad \mathcal{I}_3 = [3\frac{349}{910},\ 5\frac{419}{910}), \quad \mathcal{I}_4 = [5\frac{419}{910},\ 6.4), \\ \mathcal{I}_5 &= [6.4,\ 6\frac{43}{70}), \quad \mathcal{I}_6 = [6\frac{43}{70},\ 12\frac{349}{910}), \quad \mathcal{I}_7 = [12\frac{349}{910},\ 39]. \end{aligned}$$

There are 7 pieces of constant control in the optimal primal solution, one for each interval defined above. During interval $\mathcal{I}_i$, we apply control u^i, where

$$\begin{aligned} u^1 &= (0\ \ 4\ \ 0.2)', \quad u^2 = (0\ \ 4\ \ 0)', \quad u^3 = (2/3\ \ 4\ \ 0.2)', \quad u^4 = (4\ \ 4\ \ 0.2)', \\ u^5 &= (4\ \ 2.8\ \ 0.2)', \quad u^6 = (0\ \ 0\ \ 0.2)', \quad u^7 = (0\ \ 0\ \ 0)'. \end{aligned}$$

There are also 7 pieces of constant $\pi(t)$ in the dual optimal solution, one for each interval defined above. At interval $\mathcal{I}_i$, $\pi(t) = \pi^i$, where

$$\begin{aligned} \pi^1 &= (0.15\ \ 0\ \ 0.5)', \quad \pi^2 = (3\ \ 0\ \ 0.5)', \quad \pi^3 = (3\ \ 5\ \ 0.5)', \quad \pi^4 = (3\ \ 5\ \ -1\frac{2}{3})', \\ \pi^5 &= (3\ \ 5\ \ 0)', \quad \pi^6 = (3\ \ 5\ \ 0.5)', \quad \pi^7 = (0\ \ 0\ \ 0)', \end{aligned}$$

and $\eta(t)$ is defined from $\pi(t)$ by

$$\eta(t) = \max \{\ 0,\ -\int_t^T B'\pi(t)\ dt\ \}.$$

It is easy to check that the proposed primal and dual solutions together satisfy complementary slackness condition (4.35) and thus they are optimal for the primal and dual problem respectively. The primal solution is indeed the unique (unique almost everywhere in $[0,\ T]$) optimal solution for the problem. However, in this solution, the inventory level for workstation 3 drops to zero at time 0.5, it stays at zero for a while and becomes positive again at time $5\frac{419}{910}$, thus Structural Property 1 is violated.

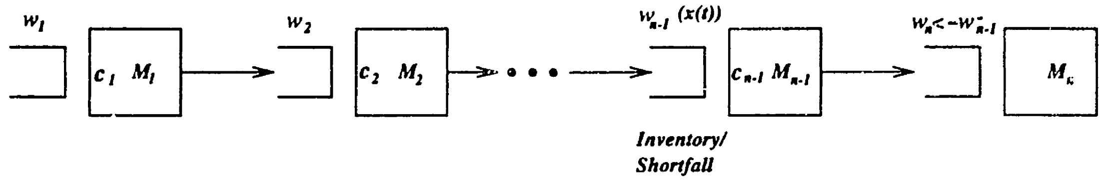

Figure 4-6: A series of queues in tandem with back ordering

4.4.2 Tandem Queues with Back Ordering

Now we consider the third problem, a generalization of problem ($LCLP1$). Negative inventory is not allowed at any workstation except workstation $n-1$. The capacity of workstation $n-1$ satisfies $c_{n-1} = \min_k c_k$. The cost for the inventory at workstation $i \leq n-2$ is $w_i \geq 0$ per unit of time, while the cost at workstation $n-1$ is $w_{n-1}^+ \geq 0$ per unit time for positive inventory and $w_{n-1}^- \geq 0$ per unit time for negative inventory. We assume $w_n < -w_{n-1}^-$. Our goal is to dynamically decide for each workstation, when to idle and how to work so as to minimize the average cost per unit time over a time interval $[0,\ T]$. We can formulate this problem as the following SCSCLP problem:

$$
\begin{aligned}
(LCLP2)\quad \textit{minimize}\quad & \int_0^T \Big(\sum_{i \neq n-1} w_i x_i(t) + w_{n-1}^+ x_{n-1}^+(t) + w_{n-1}^- x_{n-1}^-(t)\Big)\, dt \\
\textit{subject to}\quad & x(t) = x(0) + \int_0^t Bu(t)\, dt \\
& 0 \leq u(t) \leq c \\
& x_{n-1}(t) = x_{n-1}^+(t) - x_{n-1}^-(t) \\
& x_i(t) \geq 0, \qquad \text{for } i \leq n-2 \\
& x_{n-1}^+(t) \geq 0 \quad x_{n-1}^-(t) \geq 0
\end{aligned}
$$

This formulation and related assumptions are made in order to model the following mechanism in manufacturing systems. Imagine workstations from 2 to $n-1$ as $n-3$ machines and the first workstation as an infinite reservoir of raw material for these machines. If $w_1 < w_2$, the first workstation always wants to work at the maximal allowable rate c_1. When there is sufficient flow in the first workstation, the first workstation acts like a constant rate material supplier for the following machines. As shown in the following lemma, it is always profitable for workstation $n-1$ to work at its maximal allowable rate c_{n-1}. The workstation $n-1$ acts like a "virtual" buffer and workstation n acts like a customer demand that depletes this buffer at a constant rate c_{n-1}, while the buffer is replenished by the output of workstation $n-2$.

Lemma 4.9 *In any optimal solution to problem ($LCLP2$), workstation $n-1$ always works at the maximal allowable rate c_{n-1}.*

Proof The existence of a piecewise constant optimal solution to ($LCLP2$) is obvious, since the feasible control region of the problem is bounded. For any such solution to ($LCLP2$), denote t_1 as the largest time in $[0, T]$ such that the service of station $n-1$ is below rate c_{n-1}. However, a new control under which station $n-1$ works at rate c_{n-1} during $[t_1^-, t_1]$ has a better solution value in ($LCLP2$). We have a contradiction. □

Theorem 4.6 *For problem ($LCLP2$), the following structural property holds. There exist two times $0 \le t_1^* \le t_2^* \le T$. For $t \in [0, t_1^*)$, $x_{n-1}(t) < 0$; for $t \in [t_1^*, t_2^*)$, $x_i(t) \ge 0$ $\forall i \le n-1$; for $t \in [t_2^*, T)$, $x_i(t) = 0$, $\forall i \le n-2$. During each of the above three time intervals, once a workstation starts working, it will continue to work whenever possible. Once the inventory level of a workstation reaches zero, it will remain at zero level throughout the same interval. There are at most $4n-8$ number of constant pieces in the control.*

Proof Let $(u(t), x(t))$ be a basic feasible solution to problem ($LCLP2$). If $x_{n-1}(0) \ge 0$, we let $t_1^* = 0$. Otherwise, we let t_1^* be the first time such that $x_{n-1}(t)$ reaches zero (we let $t_1^* = T$ if $x_{n-1}(t)$ never reaches zero before T). We let t_2^* be the first time $x_{n-1}(t)$ drops from nonnegative to negative again (we let $t_2^* = T$ if $x_{n-1}(t)$ never drops below zero again before T).

We let

$$\mathcal{I}_1 = [0,\ t_1^*), \qquad \mathcal{I}_2 = [t_1^*,\ t_2^*), \qquad \mathcal{I}_3 = [t_2^*,\ T).$$

Obviously, $\mathcal{I}_1 = \emptyset$ if and only if $t_1^* = 0$; $\mathcal{I}_2 = \emptyset$ if and only if $t_1^* = t_2^*$; $\mathcal{I}_3 = \emptyset$ if and only if $t_2^* = T$.

During $\mathcal{I}_1$, exactly the same argument as for the problem $(LCLP)$ carries through and we arrive at the result that once a workstation starts working, it will continue to work whenever possible. Once the inventory level of a workstation reaches zero, it will remain at zero level throughout the interval $\mathcal{I}_1$.

During $\mathcal{I}_2$, all the workstations have nonnegative inventory. Exactly the same argument as for the problem $(LCLP)$ carries through and we also arrive at the result that once a workstation starts working, it will continue to work whenever possible. Once the inventory level of a workstation reaches zero, it will remain at zero level throughout the interval $\mathcal{I}_2$.

During $\mathcal{I}_3$, the only possible control for workstation $i \le n-2$ is $u_i(t) = 0$. By Lemma 4.9, workstation $n-1$ will always be busy. Thus the theorem is proved. □

The above theorem is an extension to the results in Perkins and Kumar [71], where it is further assumed that $0 = w_1 \le w_2 \le \cdots \le w_{n-2} \le w_{n-1}^+$ and $x_1(0)$ is sufficiently large.

4.4.3 Complexity Issues

We have shown in previous sections that there exist piecewise constant optimal controls to $(LCLP)$, $(LCLP1)$ and $(LCLP2)$ that have polynomial number breakpoints. For general $(FNET)$ problems, we have following computational complexity result.

Theorem 4.7 *Let (P) be a class of $(FNET)$ problems. If there exists a piecewise constant optimal control that has polynomial number breakpoints for every instance of (P), then (P) is in NP∩CO-NP.*

Proof As we have seen in Chapter 3, any piecewise constant optimal control can be obtained by solving some quadratic program whose data are bounded by some polynomial of the input data of (P). By assumption, the dimension of the quadratic program can also be chosen to be bounded by some polynomial of the input data size for (P), this proves that the problem is in NP. Let P be an optimal partition for (P). By Theorem 3.5, we see $AP^*(P)$ gives the optimal solution value for the dual problem. When $|P|$ is bounded by some polynomial of the input data size for (P), the input data size for $AP^*(P)$ is also bounded by some polynomial of the input data size for (P). This proves that (P) is also in CO-NP. □

Theorem 4.7 suggests that it is very likely there exist polynomial algorithms for problems $(LCLP)$, $(LCLP1)$ and $(LCLP2)$.

4.5 Controllability of the Fluid Networks

In this section, we are going to give necessary and sufficient conditions for the linear fluid networks $(FNET)$ to be stabilizable. More specifically, we attack the problem from system theory point of view, i.e., we study the controllability of the linear fluid models. We will investigate the ties in the stability conditions between the stochastic queueing control models and their fluid approximation counterparts. Our result generalizes a result by Moss [66] who only gave conditions for communication networks.

Now, let us give some definitions.

Definition 4.1 *We say a network is controllable if there exist control $u(t)$ and some finite time T, such that $x(t) = 0$ for all $t > T$. We say a network is totally controllable if it is controllable for all $x(0)$ such that* $\max(x_i(0)) \leq 1$.

Definition 4.2 *We say a network is weakly controllable if there exist control $u(t)$, some finite time T and some finite number M, such that $x(t) < M$ for all $t > T$.*

We remark that total controllability implies weak controllability. From now on, we let $w = e$, where e is the vector of all ones. We also assume $\max(x_i(0)) \leq 1$ in $(FNET)$.

Lemma 4.10 *The fluid network $(FNET)$ is totally controllable if there exists $\epsilon \in \Re^n$ such that the following system*

$$\begin{aligned} Bu + b &= -\epsilon \\ Du &\leq c \end{aligned}$$

$$\begin{aligned} \epsilon &> 0 \\ u &\geq 0 \end{aligned}$$

is feasible.

Proof Let $(\bar{u}, \epsilon)$ be a feasible solution to the above linear system. It is a fact that ϵ is bounded from below by some positive number that depends on B, b, D, c only. Without loss of generality, we assume $x(0) \neq 0$. Let

$$\delta = \min\{\min_{x_i(0)>0} \frac{\epsilon_i}{x_i(0) > 0}, 1\}.$$

So δ is bounded from below by some positive number that depends on B, b, D, c only.

Consider the following linear program

$$\begin{aligned} (LP1) \quad & \textit{minimize} && e'u \\ & \textit{subject to} && Bu + b \leq -\delta x(0) \qquad (4.42) \\ & && Du \leq c \\ & && u \geq 0 \end{aligned}$$

where e is the vector of all ones. It is feasible, since $\bar{u}$ is a feasible solution. It also has bounded objective value, so there exists an optimal solution u^* to $(LP1)$. We claim $Bu^* + b = -\delta x(0)$. Suppose the contrary, there exists i such that $(Bu^* + b)_i = -\delta x_i(0) - \eta$ for some $\eta > 0$. We construct a new feasible solution $\tilde{u}$ to $(LP1)$. Since $(Bu^* + b)_i \leq \eta$, there exists $r \in L_i$ such that $u_r^* > 0$. Let $\tilde{u}_r = u_r^* - \gamma$ and let $\tilde{u}_l = u_l^*$ for all $l \in L$ such that $l \neq r$, where $0 < \gamma < \min\{\eta, u_r^*\}$. We only need to show that $\tilde{u}$ satisfies (4.42) since all the other constraints of $(LP1)$ are trivially satisfied by $\tilde{u}$. Consider class j customers. If $j \neq i$, we have

$$(B\tilde{u} + b)_j \leq (Bu^* + b)_j \leq -\delta x_j(0).$$

If $j = i$, we also have

$$(B\tilde{u} + b)_j \leq (Bu^* + b)_i + \gamma \leq -\delta x_j(0) - \eta + \gamma \leq -\delta x_i(0).$$

So $\tilde{u}$ is feasible for $(LP1)$. However, $e'\tilde{u} = e'u^* - \gamma < e'u^*$, this contradicts the optimality of u^*. So u^* is feasible for the following linear system.

$$\begin{aligned} Bu + b &= -\delta x(0) \\ Du &\leq c \\ u &\geq 0 \end{aligned}$$

By same argument (just set δ to 0 in the above argument), we can show there exists $\bar{u}^*$ that is feasible for the linear system

$$\begin{aligned} Bu + b &= 0 \\ Du &\leq c \\ u &\geq 0 \end{aligned}$$

Let $t_0 = \frac{x(0)}{\delta}$. Consider the following control for $(FNET)$.

$$u(t) = \begin{cases} u^*, & \text{if } t \leq t_0, \\ \bar{u}^*, & \text{if } t > t_0. \end{cases}$$

We see $u(t)$ is a feasible control and $x(t) = 0$ for all $t \geq t_0$. □

Lemma 4.11 *The fluid network $(FNET)$ is not weakly controllable if the following system*

$$\begin{aligned} (WC) \quad \textit{minimize} \quad & 0 \\ \textit{subject to} \quad & Bu + b \leq 0 \\ & Du \leq c \\ & u \geq 0 \end{aligned}$$

is infeasible.

Proof Consider the dual of (WC),

$$(DWC) \quad \textit{maximize} \quad b'y - c'z$$

$$\begin{aligned} \textit{subject to} \quad & B'y + D'z \geq 0 \\ & y, z \geq 0 \end{aligned}$$

(DWC) is obviously feasible. By linear programming duality theory, there exists $y \geq 0$ and $z \geq 0$ such that $B'y + D'z \geq 0$ and $b'y - c'z > 0$. Multiply (4.1) by y we have

$$y'x(t) = y'x(0) + \int_0^t (y'Bu(t) + y'b)\, dt \tag{4.43}$$

Multiply (4.2) by z we have

$$z'Du \leq z'c \tag{4.44}$$

Substitute (4.44) into (4.43) we get

$$\begin{aligned} y'x(t) &\geq y'x(0) + \int_0^t (y'Bu(t) + y'b + z'Du(t) - z'c)\, dt \\ &= y'x(0) + \int_0^t ((B'y + D'z)'u(t) + (y'b - z'c))\, dt \\ &\geq y'x(0) + \int_0^t (y'b - z'c)\, dt \\ &= y'x(0) + (y'b - z'c)t. \end{aligned} \tag{4.45}$$

Clearly, this implies that the network is not weakly controllable. □

Lemma 4.12 *The fluid network $(FNET)$ is not totally controllable if the following system*

$$(TC) \quad \begin{aligned} & Bu + b = -\epsilon \\ & Du \leq c \\ & \epsilon > 0 \\ & u \geq 0 \end{aligned}$$

is infeasible.

Proof Consider the linear program (WC) and its dual (DWC) used in the proof of Lemma 4.11. If (WC) is infeasible, by Lemma 4.11, the network is not weakly controllable, so the network is not totally controllable either. If (WC) is feasible, there exists i, such that $(Bu + b)_i = 0$ for all u that is feasible for (WC) (Otherwise let u^i be a feasible solution to (WC) such that $(Bu + b)_i > 0$, we see $\frac{1}{n}\sum_i u^i$ is a feasible solution to (TC), contradicts the assumption of the lemma). Since (DWC) is always feasible, by linear programming theory, there exist u^* and $(y,\ z)$ that are optimal solutions to (WC) and (DWC) respectively, such that together they satisfy strict complementary slackness condition. So, $y \neq 0$. (4.45) is still valid, so we have

$$y'x(t) \geq y'x(0) + (y'b - z'c)t \geq y'x(0).$$

Clearly, this implies that the network can never be emptied if $x(0) = e$ and thus it is not totally controllable. □

Combining Lemmas 4.10, 4.11 and 4.12, we have the following theorem.

Theorem 4.8 *The fluid network $(FNET)$ is weakly controllable if and only if the following system is feasible*

$$\begin{aligned} Bu + b &\leq 0 \\ Du &\leq c \\ u &\geq 0; \end{aligned}$$

it is totally controllable if and only if

$$\begin{aligned} Bu + b &= -\epsilon \\ Du &\leq c \\ \epsilon &> 0 \\ u &\geq 0 \end{aligned}$$

is feasible.

Proof We only need to point out that if

$$\begin{aligned} Bu + b &\leq 0 \\ Du &\leq c \\ u &\geq 0 \end{aligned}$$

is feasible, then $u(t) = \bar{u}^*$ for all $t \geq 0$ is a feasible control that will keep the system inventory bounded (i.e., weakly controllable), where $\bar{u}^*$ is defined in the proof of Lemma 4.10. The other three directions are direct consequences of Lemmas 4.10, 4.11 and 4.12 respectively. □

When $(FNET)$ is specialized to a multiclass queueing network, we can define the following traffic intensity ρ for the network, as in Chen [22].

$$\rho \stackrel{\text{def}}{=} -DB^{-1}b.$$

Since for a multiclass queueing network, c is a vector of all ones and B^{-1} is a square nonpositive matrix, condition (TC) is essentially the same as requiring the traffic intensity at each workstation strictly less than one.

Chapter 5

Fluid Telephone Loss Networks

In this chapter, we analyze the solution structure for the problems arising in telephone loss networks. For this special class of linear optimal control problems with state feedback and constraints, we show that the problem admits piecewise constant optimal control solution when the service rates are independent of the origin and destination of the calls. Under the same assumption, we give a heuristic algorithm for the stochastic problem. We provide a closed form optimal solution for a two class single-link fluid loss network, which provides insights to both the optimal solution structure for general fluid telephone loss networks and the corresponding stochastic control problems.

5.1 Telephone Loss Networks

In this section, we examine the fluid telephone loss network that is slightly more general than the one discussed in Section 1.1.3. We are interested in the optimal solution structure for the problem.

5.1.1 Problem Formulation

As in Section 1.1.3, we have a network of locations i, for $i = 1, \cdots, n$, and $n \times (n-1)$ different links (i,j), for $i \neq j$, that connects location i to location j. Calls from location i to location j arise at rate λ_{ij} with a duration of $\frac{1}{\mu_{ij}}$. The capacity of link (i,j) is C_{ij}. For any $i \neq j$ and $k \neq j$, let $x_{ikj}(t)$ be the state variable that represents the number of calls at time t that are routed from location i to location j through location k. We use the convention that $x_{iij}(t)$ is the number of calls at time t that are routed directly from location i to location j. For any $i \neq j$ and $k \neq j$, let $u_{ikj}(t)$ be the control variable that represents the rate at which calls made at time t from location i to location j are routed through location k. We use the convention that $u_{iij}(t)$ is the rate at which calls made at time t are routed directly from location i to location j. Let w and $\bar{w}$ be two nonpositive vectors.

Let λ be the vector of λ_{ij} with $\lambda_{i_1j_1}$ appearing on top of $\lambda_{i_2j_2}$ if and only if $i_1 < i_2$ or $i_1 = i_2$ but $j_1 < j_2$. Let C be the vector of C_{ij} with $C_{i_1j_1}$ appearing on top of $C_{i_2j_2}$ if and only if $i_1 < i_2$ or $i_1 = i_2$ but $j_1 < j_2$. Let H and F be some nonnegative matrices. Let $x(t)$ be a column vector of $x_{ikj}(t)$ with $x_{i_1k_1j_1}(t)$ appearing on top of $x_{i_2k_2j_2}(t)$ if and only if $i_1 < i_2$ or $i_1 = i_2$ but $j_1 < j_2$ or $i_1 = i_2$ and $j_1 = j_2$ but $k_1 < k_2$. Let $u(t)$ be a column vector of $u_{ikj}(t)$ with $u_{i_1k_1j_1}(t)$ appearing on top of $u_{i_2k_2j_2}(t)$ if and only if $i_1 < i_2$ or $i_1 = i_2$ but $j_1 < j_2$ or $i_1 = i_2$ and $j_1 = j_2$ but $k_1 < k_2$. Consider the following problem (cf. $(TLNa)$ in Chapter 1):

$$
\begin{aligned}
(TLNa) \quad & \textit{minimize} && \int_0^T (w'u(t) + \bar{w}'x(t)) \ dt \\
& \textit{subject to} && x_{ikj}(t) = x_{ikj}(0) + \int_0^t (u_{ikj}(t) - \mu_{ij}x_{ikj}(t)) \ dt, \qquad i \neq j,\ k \neq j \\
& && Hu(t) \leq \lambda, \qquad (5.1) \\
& && Fx(t) \leq C, \qquad (5.2) \\
& && x(t),\ u(t) \geq 0, \qquad t \in [0,\ T].
\end{aligned}
$$

This formulation is more general than the one discussed in Section 1.1.3, since H and F are arbitrary nonnegative matrices (cf. (1.4) and (1.5)). In the remainder of this section, we assume that $\mu_{ij} = \mu$ for all i and j. With a slight abuse of notation, we introduce new variables $y_{ikj}(t) = x_{ikj}(t)\ e^{\mu t}$ and define $u_{ikj}(t) = e^{\mu t}\ u_{ikj}(t)$. Let $y(t)$ be a column vector of $y_{ikj}(t)$ defined in the same way as $x(t)$ is defined from $y_{ikj}(t)$ and let $u(t)$ be a column vector of $u_{ikj}(t)$ defined in the same way as $u(t)$ is defined from $u_{ikj}(t)$ in $(TLNa)$. We now have

$$(TLNb) \quad \textit{minimize} \quad \int_0^T \left(e^{-\mu t}w'u(t) + e^{-\mu t}\bar{w}'y(t)\right) \ dt \qquad (5.3)$$

$$
\begin{aligned}
\textit{subject to}\quad & y_{ikj}(t) = y_{ikj}(0) + \int_0^t u_{ikj}(t)\,dt, \qquad i \neq j,\ k \neq j \\
& Hu(t) \leq e^{\mu t}\lambda, && (5.4)\\
& Fy(t) \leq e^{\mu t}C, && (5.5)\\
& y(t),\ u(t) \geq 0, \qquad t \in [0,\ T].
\end{aligned}
$$

Let us first transform the objective function of the problem into a function of the control alone. From (5.3), we have

$$
\begin{aligned}
& \int_0^T e^{-\mu s}\bar{w}_{ikj}y_{ikj}(s)\,ds \\
= \;& -\frac{1}{\mu}\bar{w}_{ikj}e^{-\mu T}\int_0^T u_{ikj}(s)\,ds + \frac{1}{\mu}\int_0^T \bar{w}_{ikj}e^{-\mu s}u_{ikj}(s)\,ds \\
= \;& \frac{1}{\mu}\int_0^T (e^{-\mu s} - e^{-\mu T})\bar{w}_{ikj}u_{ikj}(s)\,ds
\end{aligned}
$$

So, $(TLNb)$ is equivalent to

$$
\begin{aligned}
(TLN)\quad \textit{minimize}\quad & \int_0^T \tilde{w}(t)'u(t)\,dt \\
\textit{subject to}\quad & y_{ikj}(t) = y_{ikj}(0) + \int_0^t u_{ikj}(t)\,dt, \qquad i \neq j,\ k \neq j \\
& Hu(t) \leq e^{\mu t}\lambda, \\
& Fy(t) \leq e^{\mu t}C, \\
& y(t),\ u(t) \geq 0, \qquad t \in [0,\ T],
\end{aligned}
$$

where

$$
\tilde{w}(t) = e^{-\mu t}(w + \frac{1}{\mu}\bar{w}) - \frac{1}{\mu}e^{-\mu T}\bar{w}.
$$

We remark that (TLN) is an $(SCSCLP)$. We also remark that the right hand side in the constraints of (TLN) is an exponential as opposed to piecewise linear function and Algorithm $\mathcal{A}$ in Chapter 3 does not directly apply.

5.1.2 Approximate Fluid Loss Network Problems

Let $R = \{t_0, \cdots, t_r\}$ be a partition of $[0,\ T]$. Let $a(t)$ be the piecewise linear extension of $\{a(t_0), a(t_1), \cdots, a(t_r)\}$, where

$$a(t_i) = \tilde{w}(t_0) - \sum_{j=1}^{i}(t_j - t_{j-1})e^{-\mu t_{j-1}}(\mu w + \bar{w}).$$

We let $c(t)$ be the piecewise linear extension of $\{c(t_0), c(t_1), \cdots, c(t_r)\}$, where

$$c(t_i) = C + \sum_{j=1}^{i}(t_j - t_{j-1})\mu e^{\mu t_{j-1}} C.$$

Consider the following approximation of (TLN):

$$\begin{aligned}
(TLN_R)\quad & \textit{minimize} && \sum_{i=0}^{r-1}\int_{t_i}^{t_{i+1}} a(t)'u(t)\ dt \\
& \textit{subject to} && y_{ikj}(t) = y_{ikj}(0) + \int_0^t u_{ikj}(t)\ dt, \qquad i \neq j,\ k \neq j \\
& && Hu(t) \leq e^{\mu t_m}\lambda, \qquad i \neq j,\ t \in [t_m,\ t_{m+1}) \\
& && Fy(t) \leq c(t), \\
& && y(t),\ u(t) \geq 0, \qquad t \in [0,\ T].
\end{aligned}$$

By the concavity of $\tilde{w}$ and the convexity of $e^{\mu t}\lambda$, $a(t)$ is an overestimate of $\tilde{w}(t)$ and $c(t)$ is an underestimate of $e^{\mu t}\lambda$. Furthermore, $a(t)$ and $c(t)$ converge uniformly to $\tilde{w}(t)$ and $e^{\mu t}\lambda$ as the size of the maximal length interval in partition R tends to zero. Due to these reasons, we also call (TLN_R) an approximate fluid loss network problem. (TLN_R) is also an $(SCSCLP)$. Consider the dual problem for (TLN_R) (cf. $(SCSCLP1^*)$ in Chapter 2):

$$\begin{aligned}
(TLN1_R^*)\quad & \textit{maximize} && -\int_0^T y(0)'\pi(t)\ dt - \sum_{m=1}^{r}\int_{t_m}^{t_{m+1}} e^{\mu t_m}\lambda'\eta(t)\ dt - \int_0^T c(t)'\xi(t)\ dt \\
& \textit{subject to} && a(t) - \int_t^T \pi(t)\ dt + H'\eta(t) \geq 0, \\
& && \pi(t) + F'\xi(t) \geq 0, \\
& && \eta(t) \geq 0,\ \xi(t) \geq 0,
\end{aligned}$$

By allowing δ functions in the $(TLN1_R^*)$, we have the following alternative dual problem for (TLN_R) (cf. $(SCSCLP^*)$ in Chapter 2):

$$\begin{array}{lll}(TLN_R^*) & maximize & -\int_0^T y(0)'\,d\pi(t) - \sum_{m=1}^{r} \int_{t_m}^{t_{m+1}} e^{\mu t_m}\lambda'\eta(t)\,dt - \int_0^T c(t)'\,d\xi(t) \\ & subject\ to & a(t) + \pi(t) + H'\eta(t) \geq 0, \\ & & \pi(t) + F'\xi(t) \leq 0, \\ & & \pi(t) \text{ is a bounded measurable VF function} \\ & & \xi(t) \text{ monotonic increasing and right continuous} \\ & & \quad \text{on } [0,\ T] \text{ with } \xi(T) = 0, \quad \pi(T) = 0 \\ & & \eta(t) \geq 0, \qquad \text{for } t \in [0,\ T]. \end{array}$$

5.1.3 Structural Results for (TLN_R)

In this section, we analyze the solution structure for (TLN_R). Let

$$b(t_{i-1}+) = b(t_i-) = e^{\mu t_{i-1}}\lambda,$$

and let P be a partition such that $R \subseteq P$. Rearrange the indices in R and let $b(t_{i-1}+) = b(t_i-) = b(t_{l-1}+)$ when $t_{i-1} \in [t_l,\ t_m)$, where t_l and t_m are two consecutive breakpoints in R. We can obtain the following discrete approximation of (TLN_R) (cf. $DP(P)$ in Chapter 3 and the $DP(P)$ in Pullan [77]).

$$\begin{array}{lll}DP_R(P) & minimize & \sum_{i=1}^{p}(t_i - t_{i-1})a\left(\frac{t_{i-1}+t_i}{2}+\right)'\hat{u}(t_{i-1}+) \\ & subject\ to & \hat{y}(t_0) = y(t_0), \\ & & -(t_i - t_{i-1})\hat{u}(t_{i-1}+) + \hat{y}(t_i) - \hat{y}(t_{i-1}) = 0, \quad i = 1, \cdots, p, \\ & & H\hat{u}(t_{i-1}+) \leq b(t_{i-1}+), \qquad i = 1, \cdots, p, \\ & & F\hat{y}(t_i) \leq c(t_i), \quad i = 1, \cdots, p, \\ & & \hat{u}(t_{i-1}+) \geq 0,\ \hat{y}(t_i) \geq 0, \quad i = 1, \cdots, p. \end{array}$$

(TLN_R) is an $(SCSCLP)$ with E being an identity matrix and the assumption on the problem data (cf. Assumption 2.1 in Chapter 2) is satisfied. By Theorem 3.8, we obtain the following structural result for the optimal solution for (TLN_R).

Theorem 5.1 *There exists an optimal solution $(u(t), y(t))$ to (TLN_R) where $u(t)$ is piecewise constant and $y(t)$ is piecewise linear with respect to a partition P such that $R \subseteq P$ and $t_{i-1} \neq t_i$ for all i. There is no duality gap between (TLN_R) and (TLN_R^*). Let $(\hat{u},\ \hat{y})$ be defined from $(u(t), y(t))$ at the breakpoints in P. Over each interval $[t_i,\ t_{i+1})$, $(\hat{u}(t_i+), \frac{\hat{y}(t_{i+1})-\hat{y}(t_i)}{t_{i+1}-t_i})$ is a convex combination of the extreme points of the following linear system (SYS_{J_i})*

$$
(SYS_{J_i}) \qquad \begin{aligned}
&\hat{u}(t_i+) - \frac{\hat{y}(t_{i+1}) - \hat{y}(t_i)}{t_{i+1} - t_i} = 0 \\
&H\hat{u}(t_i+) \leq b(t_i+) \\
&\left(F \frac{\hat{y}(t_{i+1}) - \hat{y}(t_i)}{t_{i+1} - t_i} \right)_{J_i} \leq \dot{c}(t_i) \\
&\hat{u}(t_i+) \geq 0,
\end{aligned}
$$

where J_i is the set of inequalities in $F\hat{y}(t_i) \leq c(t_i)$ that are binding.

Our objective is to analyze the solution structure for $(\hat{u},\ \hat{y})$, and thus characterize the optimal solution structure for $(TLNa)$. We will show that there exists an optimal control for $(TLNa)$ that is piecewise constant.

5.1.4 The Ordering of the Control Pieces

Here, we first give several structural properties for (SYS_{J_i}). Obviously, there are a finite number of systems (SYS_{J_i}) and for each such system, there are a finite number of extreme points. The total number of different extreme points for these systems are uniformly bounded by some number that is independent of R. The following lemma shows that for each extreme point $(u(t_i+), \frac{y(t_{i+1})-y(t_i)}{t_{i+1}-t_i})$ of (SYS_{J_i}), the scalar $\dot{a}(t_i+)'u(t_i+)$ is also independent of R. The scalar $\dot{a}(t_i+)'\hat{u}(t_i+)$ will be used later to show that there exists a piecewise constant optimal control for the original problem $(TLNa)$.

Lemma 5.1 *For every extreme point $(u(t_i+), \frac{y(t_{i+1})-y(t_i)}{t_{i+1}-t_i})$ of system (SYS_{J_i}), $\dot{a}(t_i+)'u(t_i+)$ is independent of R.*

Proof Assume $t_i \in [t_l,\ t_m)$, where t_l and t_m are two consecutive breakpoints in R. By Theorem 5.1 $(e^{-\mu t_l}u(t_i+),\ \frac{e^{-\mu t_l}y(t_{i+1})-e^{-\mu t_l}y(t_i)}{t_{i+1}-t_i})$ is an extreme point of the following system

$$
\begin{aligned}
(SYS1_{J_i}) \qquad & e^{-\mu t_l}u(t_i+) - \frac{e^{-\mu t_l}y(t_{i+1}) - e^{-\mu t_l}y(t_i)}{t_{i+1}-t_i} = 0 \\
& He^{-\mu t_l}u(t_i+) \le \lambda \\
& \left(F\frac{e^{-\mu t_l}y(t_{i+1}) - e^{-\mu t_l}y(t_i)}{t_{i+1}-t_i}\right)_{J_i} \le \mu C \\
& e^{-\mu t_l}u(t_i+) \ge 0.
\end{aligned}
$$

$(SYS1_{J_i})$ is obviously independent of R. The lemma now follows from the following equation:

$$\dot{a}(t_i+)'u(t_i+) = -(\mu w + \bar{w})'\left(e^{-\mu t_l}u(t_i+)\right).$$

□

The following lemma gives an order on the value $\dot{a}(t_i+)'u(t_i+)$ between adjacent intervals.

Lemma 5.2 *Let $(u(t), y(t))$ be an optimal solution to (TLN_R), where $u(t)$ is piecewise constant and $y(t)$ is piecewise linear with respect to a partition P such that $R \subseteq P$ and $t_{i-1} \neq t_i$ for all i. Let $(\hat{u},\ \hat{y})$ be defined from $(u(t), y(t))$ at the breakpoints in P. For all $i > 0$, we have*

$$\dot{a}(t_{i-1}+)'u(t_{i-1}+) \ge \dot{a}(t_i+)'u(t_i+)$$

Proof Suppose the contrary. Assume

$$\dot{a}(t_{i-1}+)'u(t_{i-1}+) < \dot{a}(t_i+)'u(t_i+),$$

and we will show a contradiction. We distinguish two cases, the first case being $i \notin R$ and the second case being $i \in R$.

If $i \notin R$, assume $t_i \in [t_l,\ t_m)$, where t_l and t_m are two consecutive breakpoints in R. We see that $t_{i-1} \in [t_l,\ t_m)$. Define a new solution $(\hat{\hat{u}},\ \hat{\hat{y}})$ in the following way:

$$\hat{\hat{u}}(t_j+) \quad = \quad \hat{u}(t_j+) \quad \text{for all } j \neq i-1,\ j \neq i$$

$$\begin{aligned}
\hat{\hat{y}}(t_j) &= \hat{y}(t_j) \quad \text{for all } j \neq i \\
\hat{\hat{u}}(t_i+) &= \hat{\hat{u}}(t_{i-1}+) = \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}\hat{u}(t_{i-1}+) + \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}}\hat{u}(t_i+) \\
\hat{\hat{y}}(t_i) &= \hat{y}(t_{i-1}) + (t_i - t_{i-1})\hat{\hat{u}}(t_{i-1}+)
\end{aligned}$$

Let $\tilde{u}(t)$ and $\tilde{y}(t)$ be the piecewise constant extension and piecewise linear extension of $\hat{\hat{u}}$ and $\hat{\hat{y}}$ respectively. It is easy to see that in $[t_{i-1}, t_{i+1})$, $\tilde{u}(t)$ takes on a constant control that moves the state $\tilde{y}(t)$ from $y(t_{i-1})$ to $y(t_{i+1})$. Hence we can easily show that $(\tilde{u}(t),\ \tilde{y}(t))$ is still a feasible solution for (TLN_R). As for the objective function, we have

$$\begin{aligned}
& V((\tilde{u}(t),\ \tilde{y}(t))) - V((u(t),\ y(t))) \\
= & \int_{t_{i-1}}^{t_{i+1}} (a(t)'\tilde{u}(t) - a(t)'u(t))\ dt \\
= & (t_i - t_{i-1})(\hat{\hat{u}}(t_i+) - \hat{u}(t_{i-1}+))'(a(t_{i-1}+) + \frac{t_i - t_{i-1}}{2}\dot{a}(t_i+)) + \\
& (t_{i+1} - t_i)(\hat{\hat{u}}(t_i+) - \hat{u}(t_i+))'(a(t_{i-1}+) + \frac{t_{i+1} + t_i - 2t_{i-1}}{2}\dot{a}(t_i+)) \\
= & \frac{(t_{i+1} - t_{i-1})^2}{2}\hat{\hat{u}}(t_i+)'\dot{a}(t_i+) - \frac{(t_i - t_{i-1})^2}{2}\hat{u}(t_{i-1}+)'\dot{a}(t_i+) - \\
& (t_{i+1} - t_i)\hat{u}(t_i+)'\frac{t_{i+1} + t_i - 2t_{i-1}}{2}\dot{a}(t_i+) \\
= & \frac{1}{2}(t_i - t_{i-1})(t_{i+1} - t_i)\dot{a}(t_i+)'(\hat{u}(t_{i-1}+) - \hat{u}(t_i+)) \\
< & 0
\end{aligned} \tag{5.6}$$

where $V((\tilde{u}(t),\ \tilde{y}(t)))$ and $V((u(t),\ y(t)))$ denote the objective value of solutions $(\tilde{u}(t),\ \tilde{y}(t))$ and $(u(t),\ y(t))$ in (TLN_R) respectively. But (5.6) contradicts the optimality of $(u(t),\ y(t))$.

If $i \in R$, assume $t_{i-1} \in [t_l,\ t_i)$, where t_l and t_i are two consecutive breakpoints in R. Now, define a new solution $(\hat{\hat{u}},\ \hat{\hat{y}})$ in the following way:

$$\begin{aligned}
\hat{\hat{u}}(t_j+) &= \hat{u}(t_j+) \quad \text{for all } j \neq i-1,\ j \neq i \\
\hat{\hat{y}}(t_j) &= \hat{y}(t_j) \quad \text{for all } j \neq i \\
\hat{\hat{u}}(t_{i-1}+) &= \frac{e^{\mu t_l}}{e^{\mu t_i}(t_{i+1} - t_i) + e^{\mu t_l}(t_i - t_{i-1})}((t_i - t_{i-1})\hat{u}(t_{i-1}+) + (t_{i+1} - t_i)\hat{u}(t_i+))
\end{aligned}$$

$$\hat{\hat{u}}(t_i+) = \frac{e^{\mu t_i}}{e^{\mu t_i}(t_{i+1}-t_i)+e^{\mu t_i}(t_i-t_{i-1})}((t_i-t_{i-1})\hat{u}(t_{i-1}+)+(t_{i+1}-t_i)\hat{u}(t_i+))$$
$$\hat{\hat{y}}(t_i) = \hat{y}(t_{i-1})+(t_i-t_{i-1})\hat{\hat{u}}(t_{i-1}+).$$

Let $\tilde{u}(t)$ and $\tilde{y}(t)$ be the piecewise constant extension and piecewise linear extension of $\hat{\hat{u}}$ and $\hat{\hat{y}}$ respectively. It is easy to see that in $[t_{i-1}, t_{i+1})$, $\tilde{u}(t)$ takes on a piecewise constant control that moves the state $\tilde{y}(t)$ from $y(t_{i-1})$ to $y(t_{i+1})$ while keeping the slope of $\tilde{y}(t)$ in proportion to that of the $c(t)$. We can again show that $(\tilde{u}(t),\ \tilde{y}(t))$ is a feasible solution for (TLN_R). As for the objective function, we have

$$\begin{aligned}
& V((\tilde{u}(t),\ \tilde{y}(t))) - V((u(t),\ y(t))) \\
= & \int_{t_{i-1}}^{t_{i+1}} (a(t)'\tilde{u}(t) - a(t)'u(t))\ dt \\
= & (t_i-t_{i-1})(\hat{\hat{u}}(t_{i-1}+)-\hat{u}(t_{i-1}+))'(a(t_{i-1}+)+\frac{t_i-t_{i-1}}{2}\dot{a}(t_{i-1}+)) + \\
& (t_{i+1}-t_i)(\hat{\hat{u}}(t_i+)-\hat{u}(t_i+))'(a(t_{i-1}+)+(t_i-t_{i-1})\dot{a}(t_{i-1}+)+\frac{t_{i+1}-t_i}{2}\dot{a}(t_i+)) \\
= & \frac{(t_i-t_{i-1})^2}{2}\hat{\hat{u}}(t_{i-1}+)'\dot{a}(t_{i-1}+) - \frac{(t_i-t_{i-1})^2}{2}\hat{u}(t_{i-1}+)'\dot{a}(t_{i-1}+) + \\
& \frac{(t_{i+1}-t_i)^2}{2}(\hat{\hat{u}}(t_i+)-\hat{u}(t_i+))'\dot{a}(t_i+) + (t_{i+1}-t_i)(\hat{\hat{u}}(t_i+) - \\
& \hat{u}(t_i+))'(t_i-t_{i-1})\dot{a}(t_{i-1}+) \\
= & \frac{1}{2}(t_i-t_{i-1})(t_{i+1}-t_i)(\dot{a}(t_{i-1}+)'\hat{u}(t_{i-1}+)-\dot{a}(t_i+)'\hat{u}(t_i+)) \\
< & 0
\end{aligned}$$

This also contradicts the optimality of $(u(t),\ y(t))$. □

An important consequence of the above proof is the following corollary.

Corollary 5.1 *Let $(u(t),\ y(t))$ be the optimal solution for (TLN_R) as in Lemma 5.2. If*

$$\dot{a}(t_{i-1}+)'u(t_{i-1}+) = \dot{a}(t_i+)'u(t_i+)$$

for the control on two adjacent intervals in $(u(t),\ y(t))$, then we can construct a new optimal solution $(\tilde{u}(t),\ \tilde{y}(t))$ for (TLN_R). Furthermore, $\tilde{u}(t_{i-1}+) = \tilde{u}(t_i+)$ if $t_i \notin R$ and $e^{-\mu t_l}\tilde{u}(t_{i-1}+) = e^{-\mu t_i}\tilde{u}(t_i+)$ if $t_i \in R$, where $t_{i-1} \in [t_l,\ t_m)$ and t_l, t_m are two consecutive breakpoints in R.

Let $(\hat{u},\ \hat{y})$ be defined as in Theorem 5.1. By Theorem 5.1, $(\hat{u}(t_i+), \frac{\hat{y}(t_{i+1})-\hat{y}(t_i)}{t_{i+1}-t_i})$ is a convex combination of the extreme points of the system (SYS_{J_i}), i.e.,

$$\begin{aligned}\hat{u}(\tilde{t}_i+\cdot) &= \sum_{j=1}^{k^{(i)}} \sigma_j^{(i)} u_j^{(i)},\\ \frac{\hat{y}(t_{i+1})-\hat{y}(t_i)}{t_{i+1}-t_i} &= \sum_{j=1}^{k^{(i)}} \sigma_j^{(i)} y_j^{(i)}\end{aligned} \tag{5.7}$$

for some $k^{(i)} \geq 1$, where $\sigma_j^{(i)} > 0$, $\sum_{j=1}^{k^{(i)}} \sigma_j^{(i)} = 1$, and $(u_j^{(i)},\ y_j^{(i)})$ are extreme points of (SYS_{J_i}). The following lemma gives a nice property that all the extreme points in (5.7) should satisfy.

Lemma 5.3 *We have*

$$\dot{a}(t_i+)' u_j^{(i)} = \dot{a}(t_i+)' u_{j+1}^{(i)}$$

for all $j \leq k^{(i)} - 1$.

Proof Suppose the contrary. WLOG, we may assume

$$\dot{a}(t_i+)' u_1^{(i)} = \max_j \dot{a}(t_i+)' u_{j+1}^{(i)}$$

Hence we have

$$\dot{a}(t_i+)' u_1^{(i)} > \dot{a}(t_i+\cdot)' u_j^{(i)},$$

for some $1 < j \leq k^{(i)}$. Otherwise the lemma would be true. Now, consider the following new solution $(\tilde{u}(t),\ \tilde{y}(t))$ for (TLN_R). $(\tilde{u}(t),\ \tilde{y}(t))$ equals $(u(t),\ y(t))$ except on $[t_i,\ t_{i+1})$. We let

$$\begin{aligned}\tilde{u}(t) &= u_1^{(i)} \quad \text{for } t \in [t_i,\ t_i + \tau(t_{i+1} - t_i)),\\ \tilde{u}(t) &= \frac{1}{1-\tau}\left((\sigma_1^{(i)} - \tau)u_1^{(i)} + \sum_{j=2}^{k^{(i)}} \sigma_j^{(i)} u_j^{(i)}\right) \quad \text{for } t \in [t_i + \tau(t_{i+1} - t_i), t_{i+1})\end{aligned}$$

and define $\tilde{y}(t)$ from $\tilde{u}(t)$ by the system dynamics, where τ is the largest scalar in $(0,\ \sigma_1^{(i)}]$ such that $F\tilde{y}(t_i+\tau(t_{i+1}-t_i)) \leq c(t_i+\tau(t_{i+1}-t_i))$. It is easily seen that such a τ exists and $(\tilde{u}(t),\ \tilde{y}(t))$ is feasible for (TLN_R). However, an analysis similar to (5.6) gives that $V((\tilde{u}(t),\ \tilde{y}(t))) - V((u(t),\ y(t))) = \frac{1}{2}\tau(1-\tau)(t_{i+1}-t_i)^2(\dot{a}(t_i+\cdot)'(\tilde{u}(t_{i+1}-) -$

$\tilde{u}(t_i+)) < 0$. This contradicts the optimality of $(u(t),\ y(t))$. □

5.1.5 Structural Result for $(TLNa)$

We can now prove the main optimal solution structural result for $(TLNa)$.

Theorem 5.2 *There exists a piecewise constant optimal control for* $(TLNa)$.

Proof Let $(u_R(t),\ y_R(t))$ be an optimal solution to (TLN_R) as in Lemma 5.2. Assume that $(\hat{v}_R(t),\ \hat{y}_R(t))$ is piecewise constant with respect to partition P_R, where $t_i < t_{i+1}$ for all i. Consider the following solution $(\bar{u}_R(t),\ \bar{y}_R(t))$ for $(TLNa)$:

$$\bar{u}_R(t) = e^{-\mu t_l} u_R(t) \text{ and } \bar{y}_R(t) = e^{-\mu t_l} u_R(t), \quad \text{for } t \in [t_l,\ t_m) \tag{5.8}$$

where t_l and t_m are two consecutive breakpoints in R. It is easy to see that $(\bar{u}(t),\ \bar{y}(t))$ is feasible for $(TLNa)$. Since, as the maximal distance of all adjacent breakpoints in R tends to zero, the objective value of (TLN_R) tends to that of (TLN), which is exactly equal to the objective value of $(TLNa)$. So every limit point of a sequence $\{(\bar{u}_{R^k}(t),\ \bar{y}_{R^k}(t))\}$ would be optimal for $(TLNa)$, where R^k is a partition of $[0,\ T]$ such that the maximal distance of all adjacent breakpoints in R^k tends to zero as k tends to infinity. It suffices to show that for any R^k, we can pick $(u^k_R(t),\ y^k_R(t))$ so that $\bar{u}_{R^k}(t)$ and $\bar{y}_{R^k}(t))$ constructed for $(TLNa)$ according to (5.8) are piecewise constant and piecewise linear respectively, and the total number of pieces for both $\bar{u}_{R^k}(t)$ and $\bar{y}_{R^k}(t))$ are uniformly bounded (We count the number of pieces in the following way. For any two adjacent intervals $[t_{i-1},\ t_i)$ and $[t_i,\ t_{i+1})$, we count them in the same piece for $\bar{u}_{R^k}(t)$ if and only if $\bar{u}_{R^k}(t)$ is constant over $[t_{i-1},\ t_{i+1})$. We define the number of pieces in $\bar{y}_{R^k}(t)$ as the number of pieces in $\bar{u}_{R^k}(t)$). In the remainder of the proof, we do just that.

Obviously, $\bar{u}_{R^k}(t)$ and $\bar{y}_{R^k}(t))$ are piecewise constant and piecewise linear respectively. Suppose we pick $(u^k_R(t),\ y^k_R(t))$ in such a way that the total number of pieces in $\bar{u}_{R^k}(t)$ and $\bar{y}_{R^k}(t))$ is minimum. Over any piece of $\bar{u}_{R^k}(t)$, by Lemma 5.3, $\dot{a}(t_i+)'u_{R^k}(t)$ is constant and equals to $\dot{a}(t_i+)'u^{(i)}_j$, where $(u^{(i)}_j,\ y^{(i)}_j)$ is an extreme point of (SYS_{J_i}) as in (5.7). By Lemma 5.2, $\dot{a}(t_i+)'u_{R^k}(t)$ can only decrease over any two adjacent pieces in $\bar{u}_{R^k}(t)$. By Corollary 5.1, we know that $\dot{a}(t_i+)'u_{R^k}(t)$ can only strictly decrease over two adjacent pieces in $\bar{u}_{R^k}(t)$; otherwise we can obtain another $(\bar{u}_R^k(t),\ \bar{y}_R^k(t))$ that is optimal for (TLN_{R^k}) and construct $(\bar{\bar{u}}_R^k(t),\ \bar{\bar{y}}_R^k(t))$ according to (5.8), with fewer pieces than $\bar{u}_{R^k}(t)$, which contradicts the way we picked $(u_R^k(t),\ y_R^k(t))$. Hence every piece for $\bar{u}_{R^k}(t)$ corresponds to a different value $\dot{a}(t_i+)'u_j^{(i)}$. Since according to Lemma 5.1, there are a finite number of $\dot{a}(t_i+)'u_j^{(i)}$ regardless of the choice of R, there is also a uniform upper bound on the number of pieces in $(u_R^k(t),\ y_R^k(t))$. This completes the proof. □

5.1.6 Algorithmic Implications

Let P be a partition of $[0,\ T]$ and $(\hat{u}(t),\ \hat{x}(t))$ be a feasible solution for $(TLNa)$ with $\hat{u}(t)$ piecewise constant with P. We have the following nonlinear program

$$
\begin{aligned}
NLP(P) \quad \min \quad & \sum_{i=1}^{p}(t_i - t_{i-1})(w + \frac{1}{\mu}\bar{w})'\hat{u}(t_{i-1}+) + \\
& \sum_{i=1}^{p}\frac{1}{\mu}\bar{w}'(\hat{x}(t_{i-1}) - \frac{1}{\mu}\hat{u}(t_{i-1}+))(1 - e^{-\mu(t_i - t_{i-1})}) \\
s.t. \quad & e^{\mu t_i}\hat{x}(t_i) - e^{\mu t_{i-1}}\hat{x}(t_{i-1}) = \frac{1}{\mu}\hat{u}(t_{i-1}+)(e^{\mu t_i} - e^{\mu t_{i-1}}), \\
& H\hat{u}(t_{i-1}+) \le \lambda, \\
& F\hat{x}(t_i) \le C, \\
& \hat{u}(t_i) \ge 0, \quad \forall i.
\end{aligned}
$$

In light of Theorem 5.2, we propose the following heuristic algorithm for solving $(TLNa)$, assuming $\mu_{ij} = \mu$.

Algorithm $\mathcal{D}$ $(H, F, w, \bar{w}, \lambda, \mu)$.

1. Get a KKT point for $NLP(P)$.

2. Remove zero length intervals.

3. Double the number of breakpoints as in Algorithm $\mathcal{A}$. Let P be the new partition and go to step 1.

We remark that in order to get a KKT point for $NLP(P)$, we can iteratively linearize both the constraints and the objective function and solve the resulting linear program for a new search direction (cf. $(SUBLP^k)$). Algorithm $\mathcal{D}$ can be used to periodically calculate a heuristic policy for the stochastic loss network.

5.2 A Single-Link Problem

In this section, we give a closed form optimal solution for a single-link fluid telephone loss network. The reason why we are interested in this problem is the following. First, the problem itself is nontrivial and has important applications in communications (see Kelly [52]) and inventory control. Second, it provides insights to the corresponding stochastic control problem. Third, we want to illustrate the structural difference between an $(SCSCLP)$ and a (CLP) that is not an $(SCSCLP)$.

5.2.1 A Single Link Telephone Loss Network

We consider an example of the problem discussed in Section 5.1. We have two types of arrivals to a single link, from node 1 to node 2, with capacity C. For class i calls, $i = 1, 2$, we have the following problem data:

	Arrival rate	Service rate	Reward rate
Class 1	λ_1	μ_1	w_1
Class 2	λ_2	μ_2	w_2

All the rates are positive. We assume that $w_2 > w_1 > 0$, and the initial number of calls for each class $x_i(0)$ satisfies: $x_1(0) > 0$, $x_2(0) > 0$ and $x_1(0) + x_2(0) < C$. We also assume that T is sufficiently large, in a sense to be specified later. We want to maximize the total reward over $[0,\ T]$, i.e.,

$$
\begin{aligned}
(STLN)\quad \textit{minimize}\quad & -\int_0^T (w_1x_1(t) + w_2x_2(t))\ dt\ - \frac{w_1}{\mu_1}x_1(T) - \frac{w_2}{\mu_2}x_2(T) \\
\textit{subject to}\quad & x_1(t) = x_1(0) + \int_0^t (u_1(t) - \mu_1 x_1(t))\ dt \\
& x_2(t) = x_2(0) + \int_0^t (u_2(t) - \mu_2 x_2(t))\ dt \\
& u_1(t) \le \lambda_1 \\
& u_2(t) \le \lambda_2
\end{aligned}
$$

$$x_1(t) + x_2(t) \le C$$
$$x(t),\ u(t) \ge 0, \qquad t \in [0,\ T].$$

We remark that $(STLN)$ is a special class of $(TLNa)$, however, we do not require that $\mu_1 = \mu_2$ in this section. By transforming the objective function into a function of the control alone and by introducing new variables $y_i(t) = e^{\mu_i t}\, x_i(t)$ and $\tilde{u}_i(t) = e^{\mu_i t}\, u_i(t)$, we have following equivalent problem

$$\begin{aligned}
(P)\quad & \textit{minimize} & & -\int_0^T \left(\frac{w_1}{\mu_1} e^{-\mu_1 t} \tilde{u}_1(t) + \frac{w_2}{\mu_2} e^{-\mu_2 t} \tilde{u}_2(t) \right) dt \\
& \textit{subject to} & & y_1(t) = x_1(0) + \int_0^t \tilde{u}_1(t)\ dt \\
& & & y_2(t) = x_2(0) + \int_0^t \tilde{u}_2(t)\ dt \\
& & & \tilde{u}_1(t) \le \lambda_1 e^{\mu_1 t} \\
& & & \tilde{u}_2(t) \le \lambda_2 e^{\mu_2 t} \\
& & & e^{-\mu_1 t} y_1(t) + e^{-\mu_2 t} y_2(t) \le C \\
& & & y(t),\ u(t) \ge 0, \qquad t \in [0,\ T].
\end{aligned}$$

We let $U = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ and $\tilde{w}(t) = -U^{-1} e^{-Ut} w$ and $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$. We have the following dual problem for (P)

$$\begin{aligned}
(D)\quad & \textit{maximize} & & -\int_0^T x(0)'\pi(t)\ dt - \int_0^T \lambda' e^{Ut}\eta(t)\ dt - \int_0^T C\xi(t)\ dt \\
& \textit{subject to} & & \tilde{w}(t) - \int_t^T \pi(t)\ dt + \eta(t) \ge 0 \\
& & & e^{Ut}\pi(t) + H'\xi(t) \ge 0 \qquad\qquad (5.9) \\
& & & \eta(t) \ge 0,\ \xi(t) \ge 0,
\end{aligned}$$

where $H = (1\ \ 1)$.

In the following, we propose a policy for $(STLN)$ and thus for (P). We construct dual solution for (D) that together with the primal solution satisfy the following complementary slackness conditions

$$\begin{aligned}
\eta_i(t)(u_i(t)-\lambda) &= 0; \\
\xi(t)(x_1(t)+x_2(t)-C) &= 0; \\
u_i(t)(\tilde{w}_i(t)-\int_t^T \pi_i(t)\,dt+\eta_i(t)) &= 0; \\
x_i(t)(e^{\mu_i t}\pi_i(t)+\xi(t)) &= 0, \qquad i=1,2.
\end{aligned}$$

Therefore this primal dual feasible solution pair is optimal by weak duality. We remark that we can replace (5.9) by $e^{Ut}\pi(t) + H'\xi(t) = 0$. This is because that $y(0) > 0$, the complementary slackness condition ensures $e^{Ut}\pi(t) + H'\xi(t) = 0$. In this section, we do not introduce an alternative dual problem. Instead, we allow the dual variables $\pi(t)$ in (D) to take on Dirac δ functions.

5.2.2 The Proposed Policy

Let Δ be defined as follows: Given that we are at time t_0, define $\tilde{t} \geq t_0$ as the solution for

$$(w_2 - w_1)e^{-\mu_1 \tilde{t}} = w_1\left(e^{-\mu_1 t_0} - e^{-\mu_1 \tilde{t}}\right).$$

Such a solution obviously exists. We let $\Delta = \tilde{t} - t_0$. It is clear that

$$w_1\left(e^{\mu_1 \Delta} - 1\right) = w_2 - w_1$$

and so $\Delta > 0$ and it is independent of t_0 and T. The interpretation for Δ is the following: if we accepted Class 1 calls for Δ extra time, we would have got the same reward as accepting Class 2 calls.

We propose the following control policy for $(STLN)$.

1. We accept Class 2 calls whenever possible.

2. We accept all Class 1 calls at time t_0 if

$$x_1(t_0) + x_2(t_0) < C$$

and

$$x_1(t_0)e^{-\mu_1\tilde{t}} + x_2(t_0)e^{-\mu_2\tilde{t}} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2\tilde{t}}) < C \tag{5.10}$$

for all $\tilde{t} \in (0, \Delta]$.

3. We accept Class 1 calls at time t_0 if

$$x_1(t_0) + x_2(t_0) = C,$$

at the largest possible rate to keep $x_1(t_0) + x_2(t_0)$ below C.

We remark that the left hand side of (5.10) is the number of calls in the link at time $t_0 + \tilde{t}$ if we let $u_1(t) = 0$ and $u_2(t) = \lambda_2$ for all $t \in [t_0, t_0 + \tilde{t}]$ (of course, by ignoring the capacity constraint). Depending on $\mu_1 = \mu_2$, $\mu_1 > \mu_2$ or $\mu_2 > \mu_1$, we have three different cases.

1. If $\mu_1 = \mu_2$, the policy is a threshold type policy (trunk reservation). Let $B = \max\{0, (\frac{\lambda_2}{\mu} - C)(e^{\mu\Delta} - 1)\}$. We accept class 1 customers if and only if there are more than B free circuits.

2. If $\mu_1 > \mu_2$, we accept class 1 customers if and only if the state $(x_1(t), x_2(t))$ is below the following line

$$x_1(t)e^{-\mu_1\Delta} + x_2(t)e^{-\mu_2\Delta} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2\Delta}) = C.$$

3. If $\mu_2 > \mu_1$, the control is more complicated. If $\frac{\lambda_2}{\mu_2} \geq C$, the policy is the same to $\mu_1 > \mu_2$, i.e., we accept class 1 customers if and only if the state $(x_1(t), x_2(t))$ is below the line

$$x_1(t)e^{-\mu_1\Delta} + x_2(t)e^{-\mu_2\Delta} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2\Delta}) = C.$$

   Otherwise, the control is characterized by a switching curve (see Figure 5-1, we accept class 1 customers only if the system state is in the shaded area). The switching curve is defined by the following three curves:

$$x_1 + x_2 = C,$$

$$x_1 e^{-\mu_1 \Delta} + x_2 e^{-\mu_2 \Delta} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2 \Delta}) = C$$

and

$$\frac{\lambda_2}{\mu_2} = x_2 + \frac{\mu_1}{\mu_2} x_1 \left(\frac{\mu_2 - \mu_1}{C\mu_2 - \lambda_2} x_1 \right)^{\frac{\mu_2 - \mu_1}{\mu_1}}$$

In the remaining of this section, we give detailed analysis for the proposed policy according to the above three cases.

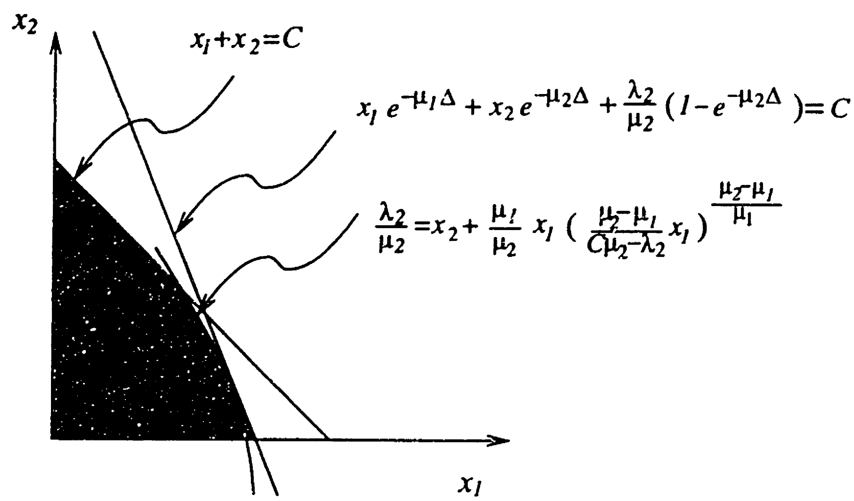

Figure 5-1: The switching curve policy

5.2.3 The Case $\mu_1 = \mu_2 = \mu$

First we consider the case $\mu_1 = \mu_2 = \mu$. The policy reduces to: always accept Class 2, accept Class 1 at time t at rate λ_1 if $x_1(t) + x_2(t) < C$ and

$$x_1(t)e^{-\mu\Delta} + x_2(t)e^{-\mu\Delta} + \frac{\lambda_2}{\mu}(1 - e^{-\mu\Delta}) < C$$

i.e.,

$$x_1(t) + x_2(t) < (C - \frac{\lambda_2}{\mu})e^{\mu\Delta} + \frac{\lambda_2}{\mu}, \tag{5.11}$$

accept Class 1 at time t at a rate to keep $x_1(t) + x_2(t) = C$ if $x_1(t) + x_2(t) = C$.

We construct dual optimal solutions for the following situations.

Case 1: If $C \geq \frac{\lambda_2}{\mu}$, (5.11) is always satisfied. We let

$$u_1(t) = \begin{cases} \lambda_1, & \text{if } x_1(t) + x_2(t) < C \\ \mu C - \lambda_2, & \text{otherwise} \end{cases}$$

and let

$$u_2(t) = \lambda_2.$$

Let t^* be the first time such that $x_1(t^*) + x_2(t^*) = C$ If no such t^* exists, we let $t^* = T$. It is easy to see that our policy is optimal and the dual optimal solution can be constructed as follows:

For $t \in [t^*,\ T]$,

$$\pi_1(t) = \pi_2(t) = -w_1 e^{-\mu t}, \quad \xi(t) = w_1, \quad \eta_1(t) = 0, \quad \eta_2(t) = (w_2 - w_1)\int_t^\infty e^{-\mu t}\, dt.$$

For $t \in [0,\ t^*)$,

$$\begin{aligned} \pi_1(t) &= \pi_2(t) = 0, \\ \xi(t) &= 0, \\ \eta_1(t) &= w_1 \int_t^{t^*} e^{-\mu t}\, dt, \quad \eta_2(t) = w_2 \int_t^{t^*} e^{-\mu t}\, dt + (w_2 - w_1)\int_{t^*}^\infty e^{-\mu t}\, dt \end{aligned}$$

Cases 2 and 3: If $C < \frac{\lambda_2}{\mu}$, it is easy to see

$$\begin{aligned} B &\stackrel{\text{def}}{=} C - ((C - \frac{\lambda_2}{\mu})e^{\mu\Delta} + \frac{\lambda_2}{\mu}) \\ &= (\frac{\lambda_2}{\mu} - C)(e^{\mu\Delta} - 1) \\ &> 0 \end{aligned}$$

Our policy gives a trunk reservation type of policy, i.e., we accept Class 1 only when

there are more than B free circuits in the link. In order to construct dual variables, we further consider the following two situations.
Case 2: $x_1(0)+x_2(0) < C-B \leq C$. Let $u_1(t)=\lambda_1$, and $u_2(t)=\lambda_2$. Since $\frac{\lambda_2}{\mu} > C$, we see that $x_1(t)+x_2(t)$ will be monotonically increasing, let t^* be the time that $x_1(t^*)+x_2(t^*)=C-B$. After t^*, we let $u_1(t)=0$ and $u_2(t)=\lambda_2$. By (5.11), we have $x_1(t^*+\Delta)+x_2(t^*+\Delta)=C$. After $t^*+\Delta$, we let $u_1(t)=0$, $u_2(t)=\mu C$.
Case 3: $x_1(0)+x_2(0) \geq C-B$. Let $u_1(t)=0$, and $u_2(t)=\lambda_2$. Since $\frac{\lambda_2}{\mu} > C$, by (5.11), there exists $t^* \leq \Delta$ such that $x_1(t^*)+x_2(t^*)=C$. After t^*, we let $u_1(t)=0$, $u_2(t)=\mu C$.
For *Case 2*, the dual variables are:

$$\begin{aligned}\pi_1(T) &= -\frac{w_2}{\mu}e^{-\mu T}\delta(T), \quad \pi_2(T) = -\frac{w_2}{\mu}e^{-\mu T}\delta(T),\\ \xi(T) &= \frac{w_2}{\mu}\delta(T), \quad \eta_1(T)=\eta_2(T)=0.\end{aligned}$$

For $t \in [t^*+\Delta,\ T)$,

$$\pi_1(t)=\pi_2(t)=-w_2e^{-\mu t}, \quad \xi(t)=w_2, \quad \eta_1(t)=\eta_2(t)=0.$$

For $t \in [t^*, t^*+\Delta)$,

$$\pi_1(t)=\pi_2(t)=0, \quad \xi(t)=0, \quad \eta_1(t)=0, \quad \eta_2(t)=w_2\int_t^{t^*+\Delta} e^{-\mu t}\,dt.$$

For $t \in [0,\ t^*)$,

$$\begin{aligned}\pi_1(t) &= \pi_2(t)=0, \quad \xi(t)=0,\\ \eta_1(t) &= w_1\int_t^{t^*} e^{-\mu t}\,dt, \quad \eta_2(t)=w_2\int_t^{t^*+\Delta} e^{-\mu t}\,dt\end{aligned}$$

For *Case 3*, the dual variables are:

$$\begin{aligned}\pi_1(T) &= -\frac{w_2}{\mu}e^{-\mu T}\delta(T), \quad \pi_2(T) = -\frac{w_2}{\mu}e^{-\mu T}\delta(T),\\ \xi(T) &= \frac{w_2}{\mu}\delta(T), \quad \eta_1(T)=\eta_2(T)=0.\end{aligned}$$

For $t \in [t^*, T)$,

$$\pi_1(t) = \pi_2(t) = -w_2 e^{-\mu t}, \quad \xi(t) = w_2, \quad \eta_1(t) = \eta_2(t) = 0.$$

For $t \in [0, t^*)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = 0, \quad \eta_2(t) = w_2 \int_t^{t^*} e^{-\mu t}\, dt$$

For both *Case 2* and *Case 3*, we require $T \geq t^* + \Delta$.

5.2.4 The Case $\mu_1 > \mu_2$

Next, we consider the case $\mu_1 > \mu_2$. Our policy reduces to: always accept Class 2, accept Class 1 at time t at rate λ_1 if $x_1(t) + x_2(t) < C$ and

$$x_1(t)e^{-\mu_1\Delta} + x_2(t)e^{-\mu_2\Delta} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2\Delta}) < C \tag{5.12}$$

accept Class 1 at time t at the maximal allowable rate (i.e., $u_1(t) \leq \lambda_1$ and $u_1(t) + u_2(t) \leq \mu_1 x_1(t) + \mu_2 x_2(t)$) if $x_1(t) + x_2(t) = C$.

We construct dual optimal solutions for the following three situations.

Case 4: If for some t^*, $x_1(t^*) + x_2(t^*) = C$ but (5.12) is not violated, $u_1(t) = \lambda_1$, and $u_2(t) = \lambda_2$ for $t \leq t^*$. At t^*, we see the derivative of the left hand side of (5.12) with respect to Δ is

$$\lambda_2 e^{-\mu_2\Delta} - \mu_1 x_1(t^*)e^{-\mu_1\Delta} - \mu_2 x_2(t^*)e^{-\mu_2\Delta}. \tag{5.13}$$

If $\lambda_2 > \mu_1 x_1(t^*) + \mu_2 x_2(t^*)$, we have that $\lambda_2 e^{-\mu_2 t} > \mu_1 x_1(t^*)e^{-\mu_1 t} + \mu_2 x_2(t^*)e^{-\mu_2 t}$ for all $t \in [0, \Delta]$ and so $x_1(t^*)e^{-\mu_1 t} + x_2(t^*)e^{-\mu_2 t} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2 t})$ is strictly increasing function of t. So,

$$x_1(t^*)e^{-\mu_1\Delta} + x_2(t^*)e^{-\mu_2\Delta} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2\Delta}) > C$$

and we have a contradiction. So $\lambda_2 \leq \mu_1 x_1(t^*) + \mu_2 x_2(t^*)$.

After t^*, we let $u_1(t) = \mu_1 x_1(t) + \mu_2 x_2(t) - \lambda_2$ and $u_2(t) = \lambda_2$, until for some $\tilde{t}$, $u_1(\tilde{t}) = \lambda_1$ (we let $\tilde{t} = T$ if no such $\tilde{t}$ exists). This can happen only when $x_1(t)$ is

strictly increasing and $x_2(t)$ is strictly decreasing for $t > t^*$. After $\tilde{t}$, $x_1(t)$ continues to increase and $x_2(t)$ continues to decrease and we let $u_1(t) = \lambda_1$ and $u_2(t) = \lambda_2$. It is easy to see that $x_1(t) + x_2(t) < C$ for all $t > \tilde{t}$. We construct a dual feasible solution as follows:
For $t \in [\tilde{t},\ T]$,

$$\pi_1(t) = \pi_2(t) = \xi(t) = 0, \quad \eta_1(t) = \int_t^\infty w_1 e^{-\mu_1 t}\, dt, \quad \eta_2(t) = \int_t^\infty w_2 e^{-\mu_2 t}\, dt.$$

For $t \in [t^*, \tilde{t})$,

$$\begin{aligned}
\pi_1(t) &= -w_1 e^{-\mu_1 t} - \delta(\tilde{t})\eta_1(\tilde{t}), \quad \pi_2(t) = -w_1 e^{-\mu_2 t} - \delta(\tilde{t})\eta_1(\tilde{t})e^{(\mu_1-\mu_2)\tilde{t}}, \\
\xi(t) &= w_1 + \delta(\tilde{t})\eta_1(\tilde{t})e^{\mu_1 \tilde{t}}, \\
\eta_1(t) &= 0, \quad \eta_2(t) = \eta_2(\tilde{t}) - \eta_1(\tilde{t})e^{(\mu_1-\mu_2)\tilde{t}} + \int_t^{\tilde{t}} (w_2 - w_1)e^{-\mu_2 t}\, dt.
\end{aligned}$$

For $t \in [0,\ t^*)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = w_1 \int_t^{t^*} e^{-\mu_1 t}\, dt, \quad \eta_2(t) = w_2 \int_t^{t^*} e^{-\mu_2 t}\, dt + \eta_2(t^*)$$

Case 5: For some $t^* > 0$, (5.12) is violated, for $t \in [t^*,\ t^* + \Delta)$, we let $u_1(t) = 0$ and $u_2(t) = \lambda_2$. It is easy to see that $x_1(t^* + \Delta) + x_2(t^* + \Delta) = C$. After $t^* + \Delta$, we let $u_1(t) = 0$ and $u_2(t) = \mu_1 x_1(t) + \mu_2 x_2(t)$. Obviously, $u_2(t)$ is decreasing after $t^* + \Delta$ and this control can be carried out until T. We construct a dual solution as follows:
For $t \in [t^* + \Delta,\ T]$,

$$\begin{aligned}
\pi_1(t) &= -w_2 e^{-\mu_1 t} + \delta(T)\tilde{w}_2(T)e^{(\mu_2-\mu_1)T}, \quad \pi_2(t) = -w_2 e^{-\mu_2 t} + \delta(T)\tilde{w}_2(T), \\
\xi(t) &= w_2 + \delta(T)\tilde{w}_2(T)e^{\mu_2 T}, \quad \eta_1(t) = \eta_2(t) = 0.
\end{aligned}$$

For $t \in [t^*, t^* + \Delta)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = 0, \quad \eta_2(t) = w_2 \int_t^{t^*+\Delta} e^{-\mu_2 t}\, dt.$$

For $t \in [0,\ t^*)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = w_1 \int_t^{t^*} e^{-\mu_1 t}\, dt, \quad \eta_2(t) = w_2 \int_t^{t^*+\Delta} e^{-\mu_2 t}\, dt$$

Case 6: At $t^* = 0$, (5.12) is violated. We see there exists $\tilde{\Delta} \in [0, \Delta]$, such that (5.12) holds with equality. For $t \in [t^*,\ t^* + \tilde{\Delta})$, we let $u_1(t) = 0$ and $u_2(t) = \lambda_2$. It is easy to see that $x_1(t^* + \tilde{\Delta}) + x_2(t^* + \tilde{\Delta}) = C$. After $t^* + \tilde{\Delta}$, we let $u_1(t) = 0$ and $u_2(t) = \mu_1 x_1(t) + \mu_2 x_2(t)$. Obviously, $u_2(t)$ is decreasing after $t^* + \tilde{\Delta}$ and this control can be carried out till T. We construct dual solution as follows:
For $t \in [t^* + \tilde{\Delta},\ T]$,

$$\begin{aligned}
\pi_1(t) &= -w_2 e^{-\mu_1 t} + \delta(T)\tilde{w}_2(T)e^{(\mu_2-\mu_1)T}, \quad \pi_2(t) = -w_2 e^{-\mu_2 t} + \delta(T)\tilde{w}_2(T), \\
\xi(t) &= w_2 + \delta(T)\tilde{w}_2(T)e^{\mu_2 T}, \quad \eta_1(t) = \eta_2(t) = 0.
\end{aligned}$$

For $t \in [t^*, t^* + \tilde{\Delta})$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = 0, \quad \eta_2(t) = w_2 \int_t^{t^*+\Delta} e^{-\mu_2 t}\, dt$$

For *Case 4*, we require $T \geq \tilde{t}$ if $u_1(\tilde{t}) = \lambda_1$. For *Cases 5* and *6*, we require that $T \geq t^* + \Delta$.

5.2.5 The Case $\mu_2 > \mu_1$

In this case, our policy reduces to: always accept Class 2, accept Class 1 at time t at rate λ_1 if $x_1(t) + x_2(t) < C$ and

$$x_1(t)e^{-\mu_1 s} + x_2(t)e^{-\mu_2 s} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2 s}) < C \tag{5.14}$$

for all $s \in [0, \Delta]$, accept Class 1 at time t at the maximal allowable rate (i.e., $u_1(t) \leq \lambda_1$ and $u_1(t) + u_2(t) \leq \mu_1 x_1(t) + \mu_2 x_2(t)$) if $x_1(t) + x_2(t) = C$.

We construct dual optimal solutions for the following four situations.
Cases 7 and *8*: For some t^*, $x_1(t) + x_2(t) = C$ and (5.14) is not violated before t^*. We see $u_1(t) = \lambda_1$ and $u_2(t) = \lambda_2$ for $t \leq t^*$. Obviously, at t^*, $\lambda_2 \leq \mu_1 x_1(t^*) + \mu_2 x_2(t^*)$. For $t \geq t^*$, let $u_1(t) = \mu_1 x_1(t) + \mu_2 x_2(t) - \lambda_2$ and $\mu_2(t) = \lambda_2$.

Case 7:$\lambda_2 > \mu_2 x_2(t^*)$. There exists $\tilde{t}$ such that $u_1(\tilde{t}) = \lambda_1$. $u_1(t)$ is monotonically increasing during $[t^*,\ \tilde{t}]$. After $\tilde{t}$, i.e., for $t \in (\tilde{t},\ T]$, we let $u_1(t) = \lambda_1$ and $u_2(t) = \lambda_2$. It is easy to see that $x_1(t) + x_2(t) < C$ for all $t > \tilde{t}$.
Case 8: $\lambda_2 \leq \mu_2 x_2(t^*)$. In this case $u_1(t)$ is monotonically decreasing in $[t^*,\ T]$, $x_1(t)$ is increasing and $x_2(t)$ is decreasing. $u_1(t)$ will always be nonnegative, since $x_2(t)$ will always be no less than $\frac{\lambda_2}{\mu_2}$. Dual variables for both cases that are complementary to primal solutions can be specified as follows:
For *Case 7*, the dual variables are:
For $t \in [\tilde{t},\ T]$,

$$\pi_1(t) = \pi_2(t) = \xi(t) = 0, \quad \eta_1(t) = \int_t^{\infty} w_1 e^{-\mu_1 t}\, dt, \quad \eta_2(t) = \int_t^{\infty} w_2 e^{-\mu_2 t}\, dt.$$

For $t \in [t^*, \tilde{t})$,

$$\begin{aligned}
\pi_1(t) &= -w_1 e^{-\mu_1 t} - \delta(\tilde{t})\eta_1(\tilde{t}), \quad \pi_2(t) = -w_1 e^{-\mu_2 t} - \delta(\tilde{t})\eta_1(\tilde{t}) e^{(\mu_1 - \mu_2)\tilde{t}}, \\
\xi(t) &= w_1 + \delta(\tilde{t})\eta_1(\tilde{t}) e^{\mu_1 \tilde{t}}, \\
\eta_1(t) &= 0, \quad \eta_2(t) = \eta_2(\tilde{t}) - \eta_1(\tilde{t}) e^{(\mu_1 - \mu_2)\tilde{t}} + \int_t^{\tilde{t}} (w_2 - w_1) e^{-\mu_2 t}\, dt.
\end{aligned}$$

For $t \in [0,\ t^*)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = w_1 \int_t^{t^*} e^{-\mu_1 t}\, dt, \quad \eta_2(t) = w_2 \int_t^{t^*} e^{-\mu_2 t}\, dt + \eta_2(t^*).$$

For *Case 7*, we require $T \geq \tilde{t}$.
For *Case 8*, the dual variables are:
For $t \in [t^*,\ T]$,

$$\begin{aligned}
\pi_1(t) &= -w_1 e^{-\mu_1 t} + \delta(T)\tilde{w}_2(T), \quad \pi_2(t) = -w_1 e^{-\mu_1 t} + \delta(T)\tilde{w}_2(T) e^{(\mu_1 - \mu_2)T}, \\
\xi(t) &= w_2 + \delta(T) e^{\mu_1 T}, \quad \eta_1(t) = 0, \quad \eta_2(t) = \int_t^T (w_2 - w_1) e^{-\mu_2 t}\, dt.
\end{aligned}$$

For $t \in [0,\ t^*)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = w_1 \int_t^{t^*} e^{-\mu_1 t}\, dt, \quad \eta_2(t) = w_2 \int_t^{t^*} e^{-\mu_2 t}\, dt + \eta_2(t^*)$$

Suppose for some t^*, (5.14) is violated, we distinguish two cases,
Case 9: $t^* = 0$, i.e.,

$$x_1(0)e^{-\mu_1 \bar{t}} + x_2(0)e^{-\mu_2 \bar{t}} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2 \bar{t}}) \geq C$$

for some $\bar{t} \in (0, \ \Delta]$ (we pick the smallest such $\bar{t}$). Let $u_1(t) = 0$ and $u_2(t) = \lambda_2$, for all $t < \bar{t}$. Obviously, $x_1(\bar{t}) + x_2(\bar{t}) = C$, $x_1(t) + x_2(t) < C$ for all $t < \bar{t}$ and $\lambda_2 \geq \mu_1 x_1(\bar{t}) + \mu_2 x_2(\bar{t})$, $x_1(t)$ decreases and $x_2(t)$ increases during $[0, \ \bar{t}]$. Let $u_1(t) = 0$ and $u_2(t) = \mu_1 x_1(t) + \mu_2 x_2(t)$ after $\bar{t}$, till $u_2(\tilde{t}) = \lambda_2$. After $\tilde{t}$, let $u_1(t) = \mu_1 x_1(t) + \mu_2 x_2(t) - \lambda_2$ and $u_2(t) = \lambda_2$, till $u_1(\tilde{\tilde{t}}) = \lambda_1$. After $\tilde{\tilde{t}}$, we let $u_1(t) = \lambda_1$ and $u_2(t) = \lambda_2$. It is a fact that $x_1(t) + x_2(t) < C$ for $t \in (\tilde{t}, \ T]$ and $x_1(t) + x_2(t)$ decreases even though $x_2(t)$ increases. If no $\tilde{t}$ or $\tilde{\tilde{t}}$ exist, we let $\tilde{t} = T$ and $\tilde{\tilde{t}} = T$ respectively.
For *Case 9*, the dual variables are:
For $t \in [\tilde{\tilde{t}}, \ T]$,

$$\pi_1(t) = \pi_2(t) = \xi(t) = 0, \quad \eta_1(t) = \int_t^\infty w_1 e^{-\mu_1 t} \, dt, \quad \eta_2(t) = \int_t^\infty w_2 e^{-\mu_2 t} \, dt.$$

For $t \in [\tilde{t}, \ \tilde{\tilde{t}})$,

$$\begin{aligned}
\pi_1(t) &= -w_1 e^{-\mu_1 t} - \delta(\tilde{\tilde{t}})\eta_1(\tilde{\tilde{t}}), \quad \pi_2(t) = -w_1 e^{-\mu_2 t} - \delta(\tilde{\tilde{t}})\eta_1(\tilde{\tilde{t}})e^{(\mu_1 - \mu_2)\tilde{\tilde{t}}}, \\
\xi(t) &= w_1 + \delta(\tilde{\tilde{t}})\eta_1(\tilde{\tilde{t}})e^{\mu_1 \tilde{\tilde{t}}}, \\
\eta_1(t) &= 0, \quad \eta_2(t) = \eta_2(\tilde{\tilde{t}}) - \eta_1(\tilde{\tilde{t}})e^{(\mu_1 - \mu_2)\tilde{\tilde{t}}} + \int_t^{\tilde{\tilde{t}}} (w_2 - w_1)e^{-\mu_2 t} \, dt.
\end{aligned}$$

For $t \in [\bar{t}, \ \tilde{t})$,

$$\begin{aligned}
\pi_1(t) &= -w_2 e^{-\mu_1 t} - \delta(\tilde{t})\eta_2(\tilde{t})e^{(\mu_2 - \mu_1)\tilde{t}}, \quad \pi_2(t) = -w_2 e^{-\mu_2 t} - \delta(\tilde{t})\eta_2(\tilde{t}), \\
\xi(t) &= w_2 + \delta(\tilde{t})\eta_2(\tilde{t})e^{\mu_2 \tilde{t}}, \\
\eta_1(t) &= 0, \quad \eta_2(t) = 0.
\end{aligned}$$

For $t \in [0, \ \bar{t})$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = 0, \quad \eta_2(t) = w_2 \int_t^{\bar{t}} e^{-\mu_2 t} \, dt + \eta_2(\bar{t})$$

Case 10: $t^* > 0$. We pick the smallest $\bar{t} \in (0,\ \Delta]$ such that

$$x_1(t^*)e^{-\mu_1\bar{t}} + x_2(t^*)e^{-\mu_2\bar{t}} + \frac{\lambda_2}{\mu_2}(1 - e^{-\mu_2\bar{t}}) = C$$

It is easy to see that

$$\lambda_2 \geq \mu_1 x_1(t^* + \bar{t}) + \mu_2 x_2(t^* + \bar{t})$$

and if $\bar{t} < \Delta$, we have

$$\lambda_2 = \mu_1 x_1(t^* + \bar{t}) + \mu_2 x_2(t^* + \bar{t}).$$

Let $u_1(t) = \lambda_1$ and $u_2(t) = \lambda_2$, for all $t \leq t^*$ and let $u_1(t) = 0$ and $u_2(t) = \lambda_2$, for all $t \in (t^*,\ t^* + \bar{t}]$. For $t > t^* + \bar{t}$, let $u_1(t) = \mu_1 x_1(t) + \mu_2 x_2(t) - \lambda_2$ and $u_2(t) = \lambda_2$. We let $\tilde{t}$ and $\tilde{\tilde{t}}$ be defined the same as in *Case 9*. We remark that if $\bar{t} < \Delta$, then $t^* + \bar{t} = \tilde{t}$. For *Case 10*, the dual variables are:
For $t \in [\tilde{\tilde{t}},\ T]$,

$$\pi_1(t) = \pi_2(t) = \xi(t) = 0, \quad \eta_1(t) = \int_t^\infty w_1 e^{-\mu_1 t}\, dt, \quad \eta_2(t) = \int_t^\infty w_2 e^{-\mu_2 t}\, dt.$$

For $t \in [\tilde{t}, \tilde{\tilde{t}})$,

$$\begin{aligned}
\pi_1(t) &= -w_1 e^{-\mu_1 t} - \delta(\tilde{\tilde{t}})\eta_1(\tilde{\tilde{t}}), \quad \pi_2(t) = -w_1 e^{-\mu_2 t} - \delta(\tilde{\tilde{t}})\eta_1(\tilde{\tilde{t}})e^{(\mu_1-\mu_2)\tilde{\tilde{t}}}, \\
\xi(t) &= w_1 + \delta(\tilde{\tilde{t}})\eta_1(\tilde{\tilde{t}})e^{\mu_1\tilde{\tilde{t}}}, \\
\eta_1(t) &= 0, \quad \eta_2(t) = \eta_2(\tilde{\tilde{t}}) - \eta_1(\tilde{\tilde{t}})e^{(\mu_1-\mu_2)\tilde{\tilde{t}}} + \int_t^{\tilde{\tilde{t}}} (w_2 - w_1)e^{-\mu_2 t}\, dt.
\end{aligned}$$

For $t \in [t^* + \bar{t},\ \tilde{t})$,

$$\begin{aligned}
\pi_1(t) &= -w_2 e^{-\mu_1 t} - \delta(\tilde{t})\eta_2(\tilde{t})e^{(\mu_2-\mu_1)\tilde{t}}, \quad \pi_2(t) = -w_2 e^{-\mu_2 t} - \delta(\tilde{t})\eta_2(\tilde{t}), \\
\xi(t) &= w_2 + \delta(\tilde{t})\eta_2(\tilde{t})e^{\mu_2\tilde{t}}, \\
\eta_1(t) &= 0, \quad \eta_2(t) = 0.
\end{aligned}$$

For $t \in [t^*,\ t^* + \bar{t})$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = 0, \quad \eta_2(t) = w_2 \int_t^{t^*+\bar{t}} e^{-\mu_2 t}\, dt + \eta_2(t^* + \bar{t}).$$

For $t \in [0,\ t^*)$,

$$\pi_1(t) = \pi_2(t) = 0, \quad \xi(t) = 0, \quad \eta_1(t) = w_1 \int_t^{t^*} e^{-\mu_1 t}\, dt, \quad \eta_2(t) = w_2 \int_t^{t^*} e^{-\mu_2 t}\, dt$$

For *Cases 9* and *10*, we require that $T \geq \bar{\bar{t}}$ if $u_1(\bar{\bar{t}}) = \lambda_1$ and $T \geq t^* + \bar{t}$ otherwise.

Remarks

There are two differences between the case where μ_{ij}s are identical and the case where they are not. The first difference is in the primal. There always exists a primal optimal solution for $(TLNa)$, whose control is piecewise linear when μ_{ij} are equal, as shown in Theorem 5.2. The control could be piecewise exponential when μ_{ij} are not equal, as in the single-link example (e.g. *Case 4*). The second difference is in the dual, as illustrated by example $(STLN)$, the dual solution for $\mu_1 = \mu_2$ can be chosen to be bounded (without δ-functions) in $[0,\ T)$, while there are cases that δ-functions must be used in $[0,\ T)$ if $\mu_1 \neq \mu_2$.

Chapter 6

Computational Results

In this chapter, we examine the computational behavior of the new algorithm. We will give several numerical examples. These examples are either standard test problems or problems arising in queueing network scheduling, communication networks and manufacturing systems. In Section 6.1, we explain the implementation details of Algorithms $\mathcal{A}$ and $\mathcal{D}$. In Section 6.2, we give several numerical examples and in Section 6.3, we give some insights to the behavior of the algorithms.

6.1 Implementations of Algorithms $\mathcal{A}$ and $\mathcal{D}$

Algorithm $\mathcal{A}$ and Algorithm $\mathcal{D}$ have been implemented and tested on a Sparc 10/41. Algorithm $\mathcal{A}$ is for solving $(SCSCLP)$ and Algorithm $\mathcal{D}$ is for solving real time scheduling of telephone loss networks. The program are written in C. We used the academic version of $LOQO$ Version 1.08 by Vanderbei [94]. We call its subroutines to solve intermediate linear programming and quadratic programming subproblems.

The implementation of Algorithm $\mathcal{A}$ consists of four modules. The input data processing module, the output module, the successive quadratic programming module and the lower bound module. The successive quadratic programming module uses Frank-Wolfe method to iteratively solve a sequence of quadratic programs, as outlined in Algorithm $\mathcal{A}$. The lower bound module uses the partition generated by the successive quadratic programming module to calculate a dual feasible solution for the problem.

The implementation of Algorithm $\mathcal{D}$ consists of two major modules. In its inner loop, the first module discretizes time, assumes the control is piecewise constant with respect to the partition and iteratively solves for better and better primal feasible solutions. It uses a gradient type method to calculate the stationary point of the intermediate nonlinear programs. In its outer loop, the first module also iteratively doubles the number of breakpoints, as outlined in Algorithm $\mathcal{D}$. The second module simulates a stochastic telephone loss network, it periodically uses the heuristic policy produced by the first module to control routing in the telephone loss network.

6.2 Numerical Examples

In this section, we give eight numerical examples. The first three examples are continuous-time network examples, as in Philpott and Craddock [74]. The fourth example is a tandem queue example. The fifth example is a re-entrant line, as in Kumar [56]. The sixth example is a communication network, as in Hajek and Ogier [42]. The seventh example is a randomly generated fully dense problem and the last example is a fluid telephone loss network example. We report computational results for all the examples, while for the eighth example, we also provide simulation results. The examples are either standard test problems or problems arising from queueing scheduling, communication networks and manufacturing systems.

6.2.1 Dynamic Network Flows

The first three examples are continuous-time network examples, as in Philpott and Craddock [74].

Example 1

The first example is a continuous-time network programming problem (a special type of SCLP) solved in Anderson and Philpott [3]. The network is shown in Figure 6-1, G is the node arc incidence matrix of the network and H is an identity matrix. The functions describing the arc costs and flow bounds are as follows:

$$\begin{aligned} &c_{(1,2)}(t) = 10 - 0.6t, \quad c_{(1,3)}(t) = 7, \quad c_{(2,3)}(t) = 6 - 0.6t, \\ &c_{(2,4)}(t) = 2 + t, \qquad c_{(3,4)}(t) = 4. \end{aligned}$$

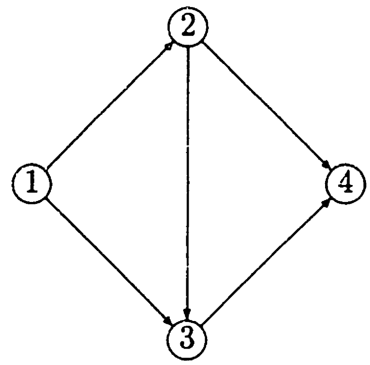

Figure 6-1: Example 1

$$b_{(1,2)}(t) = b_{(1,3)}(t) = b_{(2,4)}(t) = b_{(3,4)}(t) = 2,$$
$$b_{(2,3)}(t) = 1.$$

and $a(0) = 0$ with its derivatives defined by:

$$\dot{a}_1(t) = \begin{cases} 4, & t \in [0,5], \\ 2, & t \in (5,10], \end{cases} \qquad \dot{a}_2(t) = \begin{cases} 1, & t \in [0,5], \\ 2, & t \in (5,10], \end{cases}$$
$$\dot{a}_3(t) = \begin{cases} -1, & t \in [0,5], \\ -2, & t \in (5,10], \end{cases} \qquad \dot{a}_4(t) = \; -3, \quad t \in [0,10].$$

For this problem the optimal partition is $\{0, 3.75, 5, 8.75, 10\}$, and the optimal value is 396.25. Algorithm $\mathcal{A}$ gives the partition $\{0, 3.74996, 3.75015, 5, 8.749958, 8.75022, 10\}$ (after removing redundant intervals) and an objective value of 396.25000007. The computational sequence is shown in Table 6.1.

Example 2

The second example is posed in the network shown in Figure 6-2. Here an initial storage of 8 units must be routed from node 1 into node 4 over the interval $[0, 10]$, so $y(0) = (8, 0, 0, 0)'$, and $y(10) = (0, 0, 0, 8)'$. This problem can be formulated as an instance of SCLP problem by putting a constant demand of 1.6 per unit time at node 4 during (5, 10]. The arc costs and flow bounds are as follows:

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	410.00001	2		
1	397.50009581	2	392.49989357	0.450000
2	396.25000010	6	396.24996426	1.783333
3	396.25000007	10	396.24999701	3.683333
4	396.25000007	14	396.24999780	4.466667
5	396.25000007	14	396.24999819	5.40000

Table 6.1: Test results for Example 1.

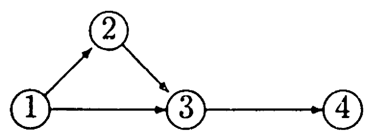

Figure 6-2: Example 2

$$c_{(1,2)}(t) = 1 + 0.6t, \quad c_{(2,3)}(t) = 1 + 1.4t,$$
$$c_{(1,3)}(t) = 12 - t, \quad c_{(3,4)}(t) = 6 - 0.2t,$$

$$b_{(1,2)}(t) = 0.6, \quad b_{(2,3)}(t) = 0.8,$$
$$b_{(1,3)}(t) = 0.8, \quad b_{(3,4)}(t) = 1.6.$$

The demand rate of each node are all zero except node 4, i.e.,

$$\dot{a}_4(t) = \begin{cases} 0, & t \in [0,5], \\ -1.6, & t \in (5,10]. \end{cases}$$

The optimal solution has the partition $\{0, 2\frac{8}{11}, 3\frac{7}{11}, 5, 10\}$, and has value of $81\frac{1}{11}$. Algorithm $\mathcal{A}$ gives the partition $\{0, 2.727274, 3.636365, 5, 10\}$ (after removing redundant intervals) and an objective value of 81.09090909. The computational sequence is shown in Table 6.2.

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	115.4959844	6		
1	81 79166783	6	80.52341959	1.133333
2	81.09090909	8	81.09091390	2.8
3	81.09090909	10	81.09090904	5.666667

Table 6.2: Test results for Example 2.

Example 3

The third problem is posed in the network shown in Figure 6-3. The arc costs and

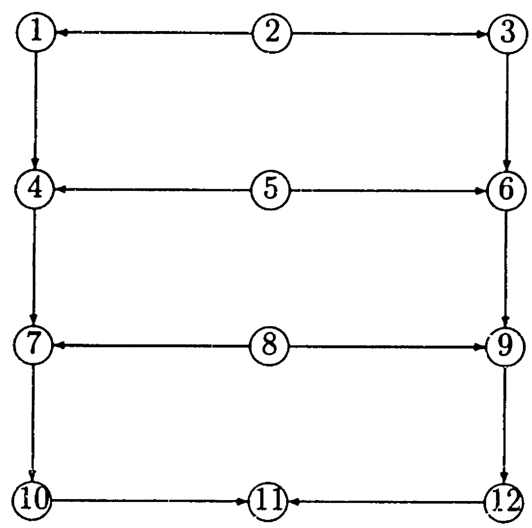

Figure 6-3: Example 3

flow bounds are as follows:

$$\begin{array}{lll}
c_{(2,1)}(t) = 7, & c_{(2,3)}(t) = 1 + 1.4t, & c_{(1,4)}(t) = 5 + 0.6t \\
c_{(3,6)}(t) = 10 - t, & c_{(5,4)}(t) = 12, & c_{(5,6)}(t) = 9, \\
c_{(4,7)}(t) = 12 - 0.6t, & c_{(6,9)}(t) = 5 + 0.6t, & c_{(8,7)}(t) = 9, \\
c_{(8,9)}(t) = 6, & c_{(7,10)}(t) = t, & c_{(9,12)}(t) = 4, \\
c_{(10,11)}(t) = 1, & c_{(12,11)}(t) = 1. &
\end{array}$$

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	728.49096903	5		
1	594.44773882	5	592.09438253	3.466667
2	593.24193904	12	593.24192962	13.65
3	593.24193557	18	593.24193392	40.633333

Table 6.3: Test results for Example 3.

$$b_{(2,1)}(t) = b_{(2,3)}(t) = 1.0, \qquad b_{(1,4)}(t) = b_{(3,6)}(t) = 2.0,$$
$$b_{(5,4)}(t) = b_{(5,6)}(t) = 1.0, \qquad b_{(4,7)}(t) = b_{(6,9)}(t) = 2.0,$$
$$b_{(8,7)}(t) = b_{(8,9)}(t) = 1.0, \qquad b_{(7,10)}(t) = b_{(9,12)}(t) = 3.0,$$
$$b_{(10,11)}(t) = b_{(12,11)}(t) = 4.0.$$

The demands are as follows:

$$\dot{a}_1(t) = \begin{cases} 0.4, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases} \qquad \dot{a}_2(t) = \begin{cases} 1.6, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases}$$
$$\dot{a}_3(t) = \begin{cases} 0.4, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases} \qquad \dot{a}_4(t) = \begin{cases} 0.4, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases}$$
$$\dot{a}_5(t) = \begin{cases} 1.6, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases} \qquad \dot{a}_6(t) = \begin{cases} 0.4, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases}$$
$$\dot{a}_7(t) = \ 0, \quad t \in [0,10], \qquad \dot{a}_8(t) = \begin{cases} 2.0, & t \in [0,5], \\ 0, & t \in (5,10], \end{cases}$$
$$\dot{a}_9(t) = \ 0, \quad t \in [0,10], \qquad \dot{a}_{10}(t) = \ 0, \quad t \in [0,10],$$
$$\dot{a}_{11}(t) = \begin{cases} 0, & t \in [0,5], \\ -6.8, & t \in (5,10], \end{cases} \qquad \dot{a}_{12}(t) = \ 0, \quad t \in [0,10].$$

The exact optimal solution for this problem is not known. Algorithm $\mathcal{A}$ gives the partition $\{0.5376, 0.5377, 1, 3.1180, 3.1183, 3.6556, 3.656, 5, 6.1290, 6.1294, 7.6774, 7.6776, 10\}$ (after removing redundant intervals) and an objective value of 593.24193557. The computational sequence is shown in Table 6.3.

The above results are comparable to the results by Philpott and Craddock [74], who used the specialized optimization code for solving network subproblems. Our

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	14284.5109	6		
1	1440.6090	6	1438.8233	6.133
2	1439.5643	13	1439.5022	26.18
3	1439.5219	27	1439.5192	121.93
4	1439.52051	39	1439.52036	295.71

Table 6.4: Test results for Example 4.

results also show that the implementation provides accurate results.

6.2.2 Multiclass Fluid Queueing Networks

Examples 4 and 5 arise from manufacturing systems.

Example 4

The fourth example is a tandem queue example as discussed in Chapter 4 (see Figure 4-3). The exogenous arrivals arrive at the first work station at rate 1. There are 25 stations. We generate all the other data randomly. The computational sequence is shown in Table 6.4.

Example 5

The fifth example is a re-entrant line, a class of fluid networks (cf. the $(FNET)$ in Chapter 4) considered in Kumar [56]. A re-entrant line is a multiclass queueing network with fixed routing. The B matrix in $(FNET)$ is the same as the node-arc incidence matrix as the for the case of tandem queues. The matrix D is a block diagonal matrix, with each block a row vector of mean service times of the customers served at the same work station. The re-entrant line we consider is shown in Figure 6-4.

We have from left to right n stations (in Figure 6-4, we have 20 stations), each station services 5 different classes of customers. There are $5n$ classes of customers in total. Class i customers will be served at Machine $\lfloor (i-1)/n + 1 \rfloor$. After Class i customer finishes service, it will become Class $i+1$ customer if $i < 5n$ and exit the system otherwise. For this system, we assume the exogenous arrival rate for Class 1 customer is 1 and zero for all other classes. We generate randomly the mean service

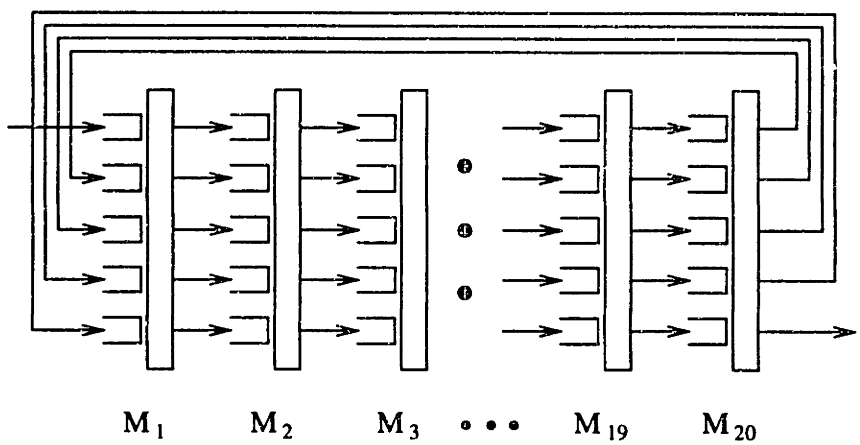

Figure 6-4: The re-entrant line for Example 5

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	20987.1355	7		
1	5986.7656	7	5956.7923	134.05
2	5965.1006	15	5962.7291	1738.2
3	5963.6674	29	5963.2700	2436.61

Table 6.5: Test results for Example 5.

time, the cost per unit time and the initial number of customers for each class of customers. The computational sequences is shown in Table 6.5.

This problem demonstrates that our algorithm can solve rather large problems.

6.2.3 Communication Networks

Example 6

The sixth problem is a communication network considered by Hajek and Ogier [42]. It is a linear fluid network problem (cf. Section 4.1). We randomly generate a graph and

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	630.4504	7		
1	2.07684	7	1.98537	428.1000
2	2.065465	15	2.061209	1738.2167
3	2.063400	27	2.062859	4754.5000
4	2.063257	43	2.063017	9165.6833

Table 6.6: Test results for Example 6.

let the coefficient matrix B in $(FNET)$ be the negative node-arc incidence matrix of the graph. The problem has 20 nodes and 100 arcs connecting the nodes. We also randomly generate the arc (link) capacities, the rate of exogenous arrivals, the initial traffic and the cost per unit time of the traffic at each node. We note that the problem is a sparse problem since the matrices B and D in $(FNET)$ are sparse matrices. The dimension of the problem is $(n_1, n_2, n_3, n_4, n_5) = (100, 20, 100, 20, 20)$. The computational sequence is shown in Table 6.6.

This problem also demonstrates that our algorithm can solve rather large problems.

6.2.4 Randomly Generated Dense SCLP Problem

Example 7

The seventh example is a randomly generated $(SCLP)$ problem. We generate all the data randomly and the coefficient matrices are fully dense matrices. The dimension of the problem is $(n_1, n_2, n_3, n_4, n_5) = (5, 10, 12, 10, 10)$. The computational sequence is shown in Table 6.7.

6.2.5 Telephone Loss Networks

Example 8

Example 8 is a telephone loss network introduced in Section 1.1.3. The underlying directed graph is shown in Figure 6-5.

The problem can be formulated as a $(TLNa)$ (by setting C_{25}, C_{52}, C_{35}, C_{53}, C_{24} and C_{42} to zero). The service rate for all the customers are 1. The arrival rate λ_{ij} is

# iter.	Obj. Value	# Pieces	Dual obj.	Time in sec.
0	85.71203	5		
1	18.85987	5	18.84785	7.2666
2	18.85617	11	18.85561	24.3833
3	18.85603	21	18.85586	63.6833
4	18.85601	29	18.85593	121.6333
5	18.85600	35	18.85594	198.9667

Table 6.7: Test results for Example 7.

50 if (i, j) is an arc in Figure 6-5. The link capacity for an arc in Figure 6-5 is 42. The reward rate for each customer served is 1 (i.e., $w_i = 0$ and $\bar{w}_i = 1$). We start the system with 8 customers in each class. For the stochastic system simulated, we assume all the arrival and service are exponentially distributed. We use Algorithm $\mathcal{D}$ to periodically get a heuristic control policy. We simulate the system for 20 unit time twenty times. We get an average reward of 542.23 per unit time. We found the reward we get is within 7 percent of an upper bound to the system[1].

6.3 Insights Gained

The above computational results show that Algorithm $\mathcal{A}$ is fairly efficient and Algorithm $\mathcal{D}$ is a very good heuristic for the stochastic system. In this section, we give some insights on why these algorithms work well.

If we fix the problem dimension, we find that the computational time grows approximately linearly with the number of control pieces allowed in the control. Figure 6-6 indicates such a relation for Example 7.

For Example 7, we plot the logarithm of the current solution value minus the best known lower bound of the objective value versus the number of breakpoints, see Figure 6-7. We find that the improvement in the objective value is sublinear. We suspect that if we had used the true objective value instead of the best know lower

[1]We thank Efthalia Chryssikou for providing a code that calculates an upper bound for the stochastic problem.

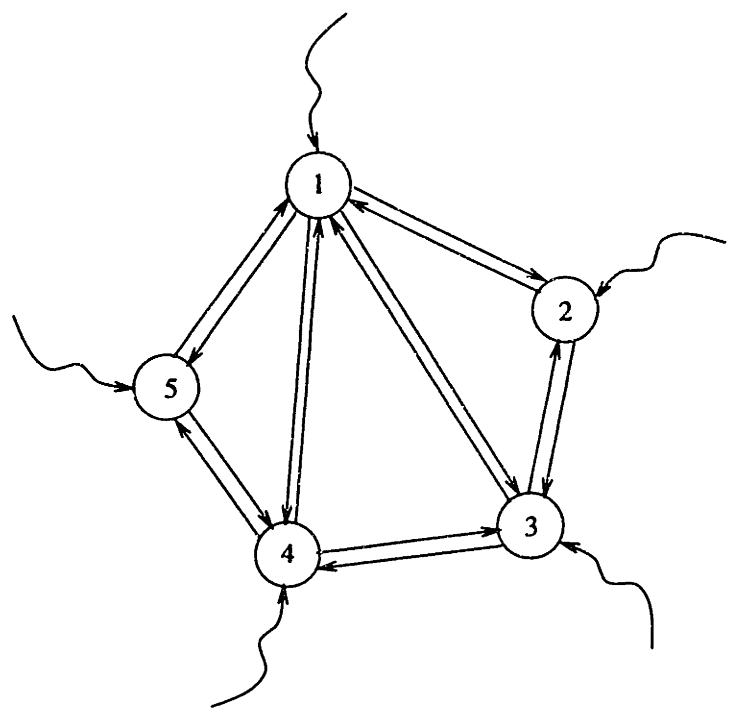

Figure 6-5: The telephone loss network for Example 8

bound, we would have get the linear rate on the improvement.

When we fix the precision requirement and vary the number of stations in Example 5, we find that the computational time grows almost quadratically with the problem dimension, as shown in Figure 6-8. This is due to the fact that the number of control pieces grows almost linearly with the problem dimension and the total number of nonzero elements in the intermediate problems grows almost quadratically with the number of stations.

It is our experience that $(SCSCLP)$ is easier to be approximated than to be exactly solved. In Figure 6-9, we plot the computational time versus the number of significant digits for Example 7. The data for Figure 6-9 is shown in Table 6.8. As shown in Figure 6-9, the computational time grows exponentially with the accuracy requirement (when we fixed the problem dimension at $(5, 10, 12, 10, 10)$). A

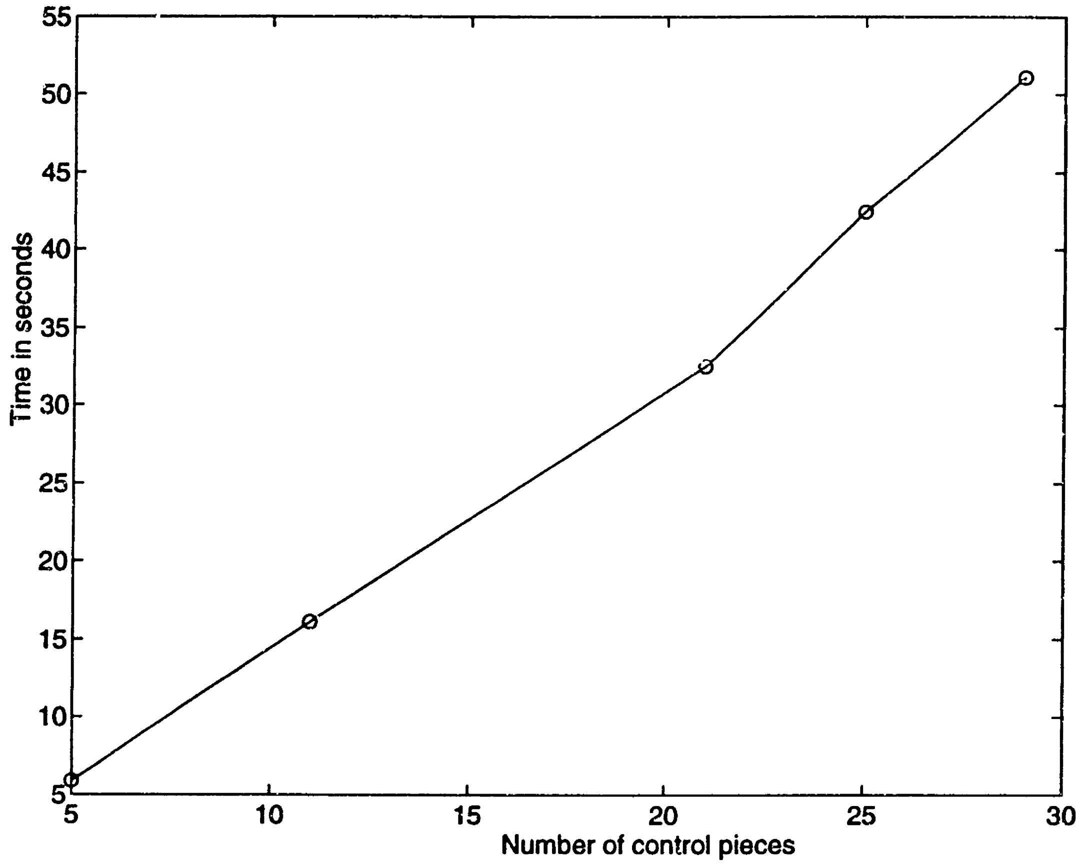

Figure 6-6: Computation time versus number of control pieces

key feature of Algorithm $\mathcal{A}$ is that it keeps the number of breakpoints as small as possible. This keeps the size of intermediate quadratic programming subproblems small. It plays a much more significant role than the nonlinearity introduced. We believe that Algorithm $\mathcal{A}$ can be made even more efficient if the special structure of the intermediate quadratic programs is exploited.

In Example 8, our heuristic behaves better when the capacity of the the links is increased. This agrees with the intuition that the larger the capacity, the smaller the influence of the randomness, and the closer the system is to a fluid model.

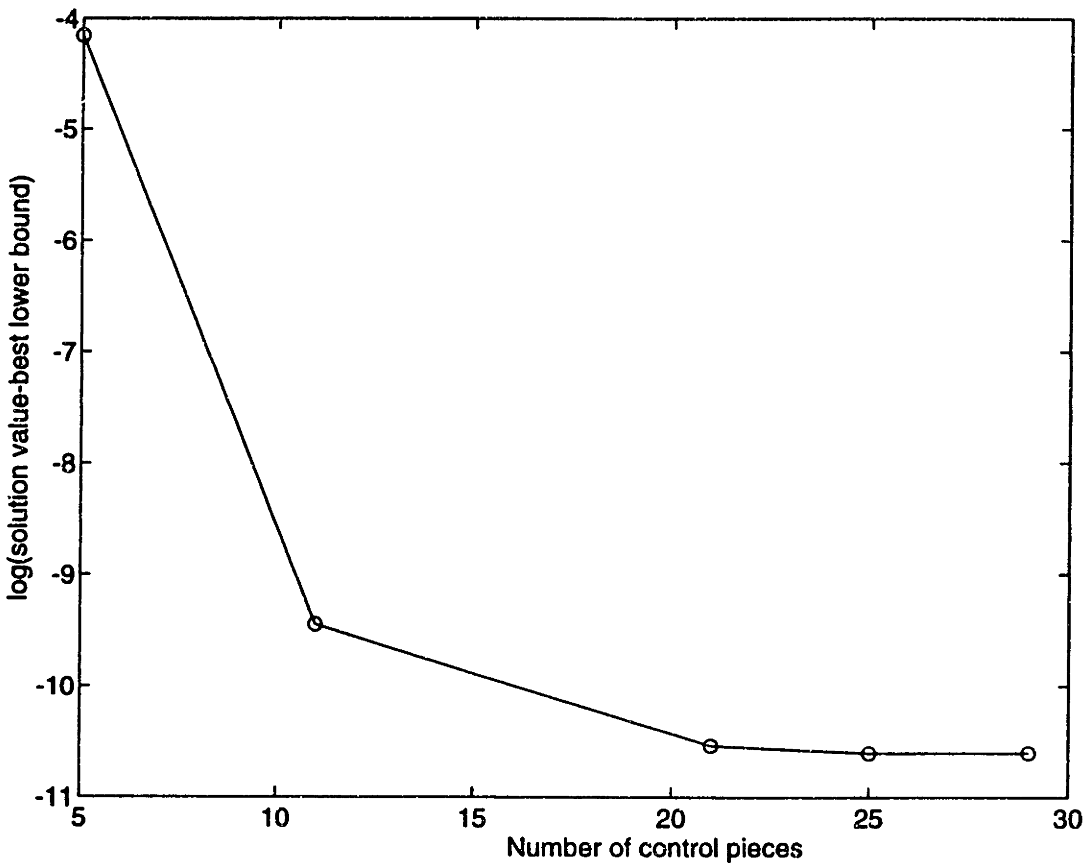

Figure 6-7: Logarithm of the current duality gap versus number of control pieces

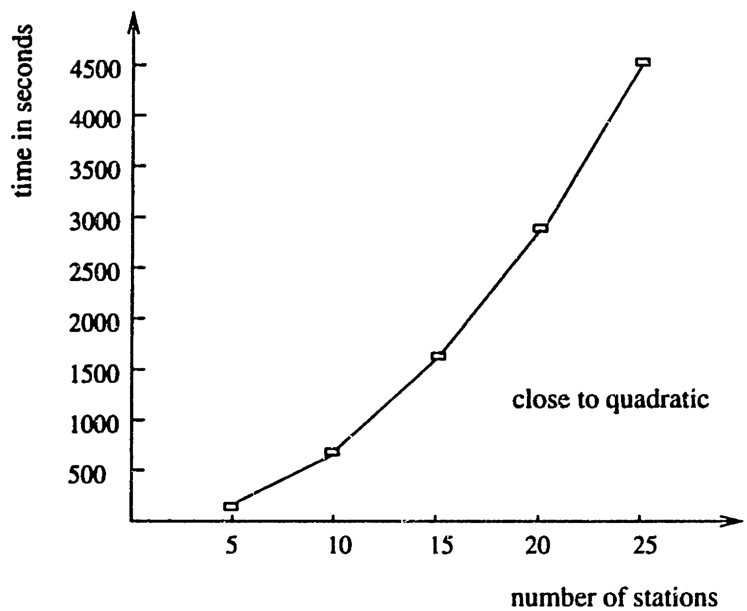

Figure 6-8: Computation time versus the number of stations
(with precision fixed at .0001)

# iter.	Inner loop	Obj. Value	# Pieces	Dual obj.	Time in sec.
1	1	85.57861839	6		1.000000
	2	33.84362951	6		1.400000
	3	18.92708484	6		1.833333
	4	18.91192386	6		2.266667
	5	18.88555273	6		2.716667
	6	18.87046051	6		3.166667
	7	18.86643960	6		3.616667
	8	18.86445830	6		4.066667
	9	18.86345834	6		4.516667
	10	18.86270212	6		4.950000
	11	18.86216700	6		5.416667
	12	18.86173743	6		5.833333
	13	18.86140729	6		6.316667
	14	18.86113686	6	18.85388267	6.733333
2	1	18.86113686	13		8.683333
	2	18.86074122	13		9.950000
	3	18.85947549	13		11.116667
	4	18.85791165	13		12.266667
	5	18.85747837	13		13.383333
	6	18.85730409	13		14.533333
	7	18.85717199	13		15.683333
	8	18.85706754	13		16.833333
	9	18.85698251	13		18.000000
	10	18.85691135	13		19.200000
	11	18.85685226	13		20.366667
	12	18.85680156	13		21.566667
	13	18.85675868	13		22.733333
	14	18.85672113	13	18.85528458	23.933333
3	1	18.85672113	27		28.333333
	2	18.85666975	27		31.166667
	3	18.85661344	27		34.066667
	4	18.85656814	27		36.766667
	5	18.85653049	27		39.733333
	6	18.85649836	27		42.416667
	7	18.85647057	27		45.383333
	8	18.85644651	27		48.083333
	9	18.85642533	27		51.166667

Table 6.8: Data for Figure 6-9.

# iter.	Inner loop	Obj. Value	# Pieces	Dual obj.	Time in sec.
	10	18.85640677	27		53.866667
	11	18.85639021	27		56.833333
	12	18.85637555	27		59.433333
	13	18.85636232	27		62.416667
	14	18.85635051	27	18.85569451	65.016667
4	1	18.85635735	37		73.283333
	2	18.85634586	37		78.033333
	3	18.85633585	37		82.000000
	4	18.85632565	37		86.250000
	5	18.85631110	37		90.100000
	6	18.85630164	37		94.466667
	7	18.85629322	37		98.333333
	8	18.85628550	37		102.566667
	9	18.85627840	37		106.433333
	10	18.85627177	37		110.666667
	11	18.85626587	37		114.516667
	12	18.85626042	37		118.883333
	13	18.85625559	37		122.733333
	14	18.85625112	37	18.85584232	126.966667
5	1	18.85625429	41		137.933333
	2	18.85625043	41		142.366667
	3	18.85624607	41		147.516667
	4	18.85624242	41		151.783333
	5	18.85623675	41		156.650000
	6	18.85623209	41		161.066667
	7	18.85622788	41		165.933333
	8	18.85622409	41		170.533333
	9	18.85622060	41		175.266667
	10	18.85621746	41		179.850000
	11	18.85621456	41		184.733333
	12	18.85621182	41		189.150000
	13	18.85620923	41		194.150000
	14	18.85620674	41	18.85587999	198.416667

Table 6.9: Data for Figure 6-9 (cont.).

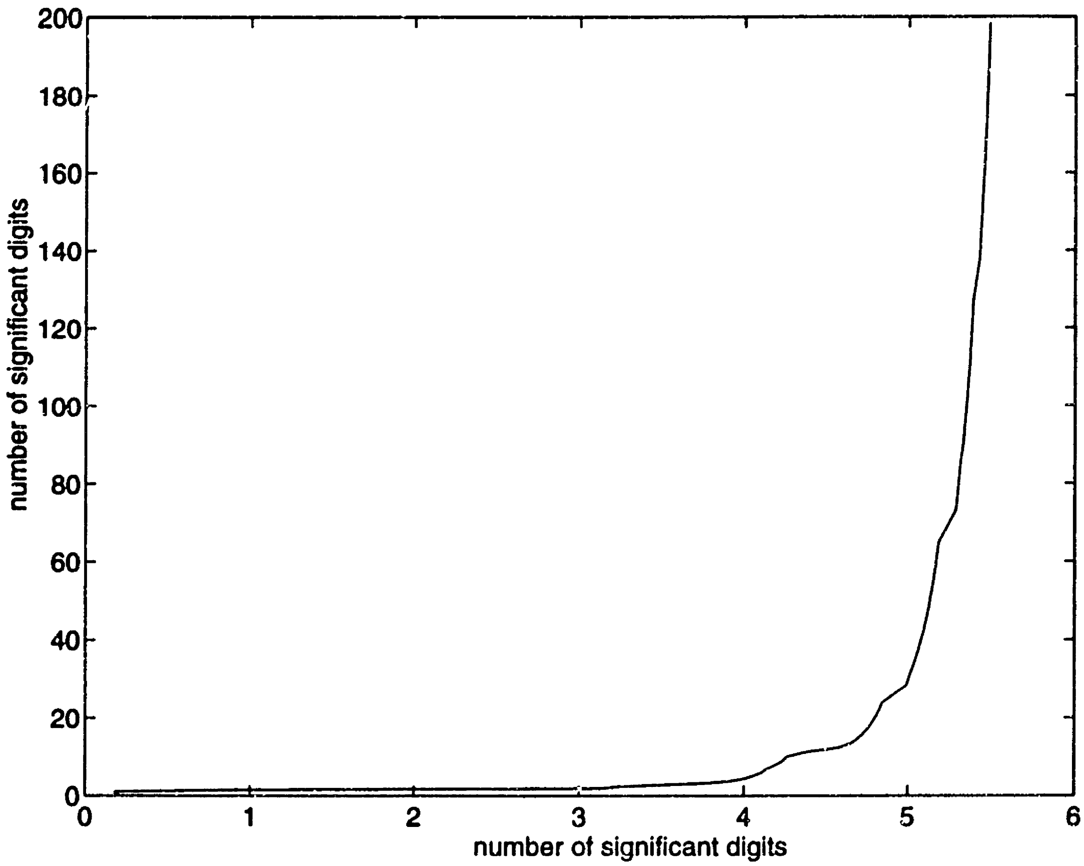

Figure 6-9: Computation time versus number of significant digits

Chapter 7

Conclusions and Open Questions

In this chapter, we summarize the main results of this thesis and point out some open questions for future study.

7.1 Summary of the Thesis

Motivated by the problems arising in queueing networks, manufacturing systems and communication networks, we proposed a larger subclass of continuous linear programming problems, i.e., SCSCLP. These problems describe time-dependent averages in the system and can be used for the control of such system in a nonstationary environment. Most importantly, these problems can be efficiently solved using mathematical programming techniques, in contrast to the traditional diffusion control approach.

As in finite dimensional linear programming, we investigated the SCSCLP problem with the help of its dual problem. We gave an alternative dual problem for SCSCLP. We developed a new algorithm called the Successive Quadratic Programming method for general SCSCLP problems under Assumption 2.1.

The new algorithm discretizes the problem over time. But unlike the other algorithms, it varies the discretization and the control simultaneously. Based on the number of constant pieces allowed in the control, we developed a quadratic program with polyhedral constraints. Even though the quadratic program is generally not convex, we applied nonlinear programming techniques such as the Frank-Wolfe method and the Matrix Splitting algorithm to get a KKT point for the quadratic program.

By gradually increasing (and occasionally decreasing) the number of pieces allowed in the control, we can get better and better approximations. We can improve any feasible solution that is not globally optimal for the SCSCLP. By bounding the size of the quadratic programming problems we encounter, we proved the finite convergence of the new algorithm. We also derived the optimal solution structure and the absence of a duality gap results as the byproducts of the new algorithm. These type of results (i.e., finite convergence, optimal solution structure and absence of a duality gap) were only known under much more restrictive assumptions.

We then developed a general framework for the fluid approximation of the multiclass queueing networks with routing. We formulated them as SCLP, a subclass of SCSCLP. We applied continuous linear programming theory to some simple queueing control problems, which includes the Klimov's problem, the single multiclass queueing control problem with separable quadratic cost, the single class tandem queueing control problem. For the first problem, we gave an index rule policy (which shows the problem is solvable in polynomial time) that solves both the fluid problem and the stochastic problem. For the second problem, we proposed a dynamic index rule that solves the fluid control problem. For the third problem, we proved the existence of a polynomial size optimal solution for the problem, which shows the problem is in NP∩CO-NP, a strong indication of the existence of a polynomial time algorithm for the problem. For the fluid multiclass queueing networks with routing, we gave simple necessary and sufficient condition for the network to be stabilizable.

We also applied continuous linear programming theory to the fluid telephone loss network problems. For this special class of linear optimal control problem with state feedback and constraints, we showed that the problem admits piecewise constant optimal control solution, when the service rates are independent of the origin and destination of the calls. This new structural result gives a heuristic for the stochastic problem. We gave a closed form optimal solution for a two class single-link fluid loss network, which provides insights to both the optimal solution structure for general fluid telephone loss networks and the corresponding stochastic control problem.

We have implemented our algorithms using programming language C. We tested our new algorithms using standard test problems and problems from manufacturing systems, communication networks and telephone loss networks. Our computational results show that our new algorithms are quite promising.

7.2 Some Open Questions

In this section, we pose two open questions.

The first open question is to characterize the optimal solution structure of the linear optimal control problem with state feedback and constraints.

The second open question is to find out the computational complexity for SCSCLP, SCLP or $(FNET)$.

Appendix A

Mathematical Background

In this appendix, we develop some mathematical background that is needed for the thesis. We give some basics on real and functional analysis. More material could be found in standard textbooks such as Rudin [83], Royden [82], Reed and Simon [80], Anderson and Nash [2] and Harrison [44].

We say that a function $f : [a,\ b] \mapsto \Re$ is measurable if $f^{-1}[(a_1, b_1)]$ is a Borel set (a set generated by a countable union of intervals) for all $a \leq a_1 < b_1 \leq b$. Let $f : [a,\ b] \mapsto \Re$ be any given measurable function. The function f is said to be absolutely continuous on $[a,\ b]$ if, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \epsilon$$

for every finite collection of nonoverlapping intervals $\{(a_i, b_i) : i = 1, \cdots, n\}$ with $a \leq a_i < b_i \leq b$ and

$$\sum_{i=1}^{n} (b_i - a_i) < \delta.$$

The following proposition is proved in Royden [82]:

Proposition A.1 *$f : [a,\ b] \mapsto \Re$ is absolutely continuous if and only if there is a measurable function $g : [a,\ b] \mapsto \Re$ such that $f(t) = f(a) + \int_a^t g(t)\ dt$ (Lebesgue integral) for all $t \in [a,\ b]$.*

The function $g(t)$ appearing in the above proposition is called a density of $f(t)$; it is not unique, but any two densities must be equal except on a set of Lebesgue measure zero. Any absolutely continuous function is differentiable almost everywhere and the derivative is a density.

We say a function $f : [a,\ b] \mapsto \Re$ is a bounded measurable function if and only if it is measurable and there exists a constant M, such that $|f(t)| < M$ a.e. on $[a,\ b]$. We denote by $L^n_\infty[a,\ b]$ the space of n dimensional vectors whose elements are bounded measurable functions over $[a,\ b]$. We denote by $L^n_1[a,\ b]$ the space of n dimensional vectors whose elements are Lebesgue integrable functions over $[a,\ b]$.

Again, given $f : [a,\ b] \mapsto \Re$. We say $f(t)$ is a function of bounded variation (or equivalent, a VF function) if the following holds:

$$\mathcal{V}(f) \stackrel{\text{def}}{=} \sup\left\{\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|\right\} < \infty,$$

where the supremum is taken over all finite partitions $a = t_0 < t_1 < \cdots < t_n = b$. We call $\mathcal{V}(f)$ the total variation of $f(t)$ over $[a,\ b]$. An important fact about a VF function is the following, see also Harrison [44]

Proposition A.2 *If $f : [a,\ b] \mapsto \Re$ and $g : [a,\ b] \mapsto \Re$ are continuous and VF functions respectively, then the Lebesgue-Stieltjes integrals $\int f(t)\ dg(t)$ and $\int g(t)\ df(t)$ both exist. Furthermore, integration by parts holds, i.e.,*

$$\int_a^b f(t)\ dg(t) = [f(b)g(b) - f(a)g(a)] - \int_a^b g(t)\ df(t)$$

We denote $BV^n[a,\ b]$ as the space of n dimensional vectors whose components are VF functions over $[a,\ b]$.

We say a function $f : [a,\ b] \mapsto \Re$ is analytic over a neighborhood of $[a,\ b]$ (or $[a,\ b)$) if there exists an $\epsilon > 0$ and an analytic function $g : (a - \epsilon,\ b + \epsilon) \mapsto \Re$ such that $f(t) = g(t)$ for all $t \in [a,\ b]$ (respectively $[a,\ b)$).

In the thesis, we need to use one particular generalized function, namely, the δ-functions. There are many ways to define such a function. Here, we take an engineering approach, for ease of understanding. Let $\mathcal{H}(t)$ be a step function (or equivalently, a Heaviside function) defined to be equal to zero for negative value t

and to unity for positive value t, i.e.,

$$\mathcal{H}(t) = \begin{cases} 0, & \text{if } t < 0 \\ 1, & \text{if } t \geq 0 \end{cases}$$

$\mathcal{H}(t)$ is not a continuous function, but it is a VF function. If the jump at the step function is at a instead of 0, it is written as $\mathcal{H}(t-a)$. We define the Dirac δ function $\delta(t)$ to be the density of $\mathcal{H}(t)$ in the following sense

$$\delta(t) = 0 \quad \text{for} \quad t \neq 0, \quad \int_a^b \delta(t)\, dt = 1 \;\; \text{only if} \;\; a < 0 < b$$

We remark that the $\delta(t)$ defined above is meaningful only when it appears under integrals. This way, we need not formally (and tediously) introduce generalized functions in the thesis. We also remark, even though that $\delta(t)$ is not an ordinary function, it can be regarded as the limit of a sequence of functions, as we can see in Chapter 3.

Let X be a linear vector space. We call a subset $\mathcal{P}$ of X a convex cone in X if it is closed under addition and multiplication by positive scalars. Let X be partially ordered by the relation $\leq$, defined by

$$x \leq y \quad \text{if } y - x \in \mathcal{P} \quad \text{and} \quad x, y \in X.$$

We write $y \geq x$ if $x \leq y$. In any vector space, we still write the null vector as 0. Clearly, $x \in \mathcal{P}$ if and only if $x \geq 0$, so we call such an x positive. $\mathcal{P}$ is called the positive cone in X.

Let X and Y be any two linear vector spaces. A bilinear form defined on $X \times Y$ is a function from $X \times Y$ to $\Re$, which we write as $\langle \cdot, \cdot \rangle$, such that $\langle x, y \rangle$ is a linear function of x for each fixed y in Y, and a linear function of y for each fixed x in X. We call (X, Y) a dual pair of spaces if X and Y have a bilinear form defined on them. Let $\mathcal{P}$ be a positive cone in X, we can define the dual positive cone of $\mathcal{P}$ in Y by the following

$$\mathcal{P}^* \stackrel{\text{def}}{=} \{y \in Y : \; \langle x, y \rangle \; \geq 0, \quad \text{for all } x \in X\}$$

Bibliography

[1] E. J. Anderson, P. Nash, and A . F. Perold, "Some Properties of a Class of Continuous Linear Programs," *SIAM J. Control Optim.*, Vol. 21, pp. 758–765, 1983.

[2] E. J. Anderson and P. Nash, *Linear Programming in Infinite Dimensional Spaces,* Wiley–interscience, Chichester, 1987.

[3] E. J. Anderson and A. B. Philpott, "A Continuous–Time Network Simplex Algorithm," *Networks*, Vol. 19, pp. 395–425, 1989.

[4] E. J. Anderson, "A Continuous Model for Job–Shop Scheduling," *Ph. D. thesis*, University of Cambridge, Cambridge, U.K., 1978.

[5] E. J. Anderson, "Basic Solutions and a 'Simplex' Method for a Class of Continuous Linear Programs," *Proceedings of 9th IFIP conference on optimization techniques,* Vol. 2, pp. 26–35, 1980.

[6] E. J. Anderson, "A New Continuous Model for Job–Shop Scheduling," *International J. Systems Science*, Vol. 12, pp. 1469–1475, 1981.

[7] K. M. Anstreicher, "Generation of Feasible Descent Directions in Continuous Time Linear Programming," *tech. report*, SOL 83–18, Department of Operations Research, Stanford University, Stanford, CA, 1983.

[8] S. Bai and S. Gershwin, "Scheduling Manufacturing Systems with Work–In–Process Inventory," *Proc. of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, pp. 557–564, 1991.

[9] S. Bai, "Scheduling Manufacturing Systems with Work–In–Process Inventory Control," *Ph.D. Thesis*, MIT Operations Research Center, Cambridge MA, 1991.

[10] R. Bellman, "Bottleneck Problem and Dynamic Programming," *Proc. Nat. Acad. Sci.*, U.S.A., Vol. 39, pp. 947–951, 1953.

[11] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, N.J., 1957.

[12] D. P. Bertsekas and P. Tseng, "Relaxation Methods for Minimum Cost Network Flow Problems," *Operations Research*, Vol. 36, pp. 93–114, 1988.

[13] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice–Hall, Englewood Cliffs, N.J., 1989.

[14] D. P. Bertsekas, *Dynamic Programming*, Academic Press, 1987.

[15] D. J. Bertsimas and J. Nino–Mora, "Conservation Laws, Extended Polymatroids and Multi–Armed Bandit Problems: a Unified Approach to Indexable Systems," *tech. rep.*, Alfred P. Sloan School of Management, MIT, 1993.

[16] Dimitris Bertsimas, Ioannis Ch. Paschalidis and John Tsitsiklis, "Scheduling of Multi–Class Queueing Network: Bounds and Achievable Performance (Extended Abstract)," *Sloan School of Management Working Paper*, Massachusetts Institute of Technology, Cambridge, MA 02139.

[17] R. N. Buie and J. Abraham, "Numerical Solutions in Continuous Linear Programming Problems," *Z. Oper. Res.*, Vol. 17, pp. 107–117, 1973.

[18] John A. Buzacott and J. George Shanthikumar, *Stochastic Models of Manufacturing Systems*, Prentice–Hall, Englewood Cliffs, N.J., 1993.

[19] H. Chen and A. Mandelbaum, "Discrete Flow Networks: Bottleneck Analysis and Fluid Approximations," submitted for publication, 1987.

[20] H. Chen and D. D. Yao, "Dynamic Scheduling of a Multiclass Fluid Network," *Operations Research*, Vol. 41, pp. 1104–1115, November–December 1993.

[21] H. Chen, "A Heterogeneous Fluid Model and Its Applications in Multiclass Queueing Networks," preprint, 1988.

[22] H. Chen, "Fluid Approximations and Stability of Multiclass Queueing Networks I: Work–Conserving Disciplines," *tech. report*, University of British Columbia, Vancouver, CA, 1994.

[23] Philippe G. Ciarlet, *Introduction to Numerical Linear Algebra and Optimization*, © 1988 Cambridge University Press. ISBN 0 521 32788.

[24] J. G. Dai and J. M. Harrison, "Steady State Analysis of RBM in Rectangle: Numerical Methods and a Queueing Application," *The Annals of Applied Probability*, Vol. 1, No. 1, pp. 16–35, 1991.

[25] Jim G. Dai and Sean Meyn, "Stability and Convergence of Moments for Multiclass Queueing Networks via Fluid Limit Models," *tech. reports,* School of Industrial and Systems Engineering and School of Mathematics, Atlanta, GA, April 1994.

[26] Jim G. Dai and Y. Wang, "Nonexistence of Brownian Models of Certain Multiclass Queueing Networks," *Queueing Systems: Theory and Applications*, Vol. 13, pp. 41–46, 1993.

[27] Jim G. Dai, "On the Positive Harris Recurrence for Multiclass Queueing Network: A Unified Approach via Fluid Limit Models," *tech. reports,* School of Industrial and Systems Engineering and School of Mathematics, Atlanta, GA, Oct, 1993.

[28] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963.

[29] W. P. Drews, "A Simplex–like Algorithm for Continuous–Time Linear Optimal Control Problems," *Optimization Methods for Resource Allocation*, R. W. Cottle and J. Krarup, eds., Crane Russak and Co. Inc., New York, pp. 309–322, 1974.

[30] Jonathan Eckstein, "Splitting Methods for Monotone Operators with Applications to Parallel Optimization," MIT thesis, 1989.

[31] Kaan Egilmez and Ali Sharifnia, "Optimal Control of a Manufacturing System Based on a Novel Continuous–Flow Model with Minimum WIP Requirement," working paper, Department of Manufacturing Engineering, Boston University, Boston, MA 02215.

[32] J. Filipiak, *Modeling and Control of Dynamic Flows in Communication Networks*, Springer–Verlag, Berlin, 1988.

[33] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to Theory of NP–Completeness*, Freeman, San Francisco, 1979.

[34] R. J. Gibbens and P. C. Reichl, "Performance Bounds Applied to Loss Networks," working paper 93–30, Statistics Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England, UK.

[35] G. H. Golub and C. F. Van Loan, *Matrix Computations*, © 1989 The Johns Hopkins University Press, Baltimore and London. ISBN 0-8018-3732-3.

[36] D. H. Griffel, *Applied Functional Analysis,* John Wiley & Sons, 1981.

[37] R. C. Grinold, "Continuous Programming," Ph. D. Thesis, Operations Research Center, University of California, Berkeley, CA, 1968.

[38] R. C. Grinold, "Continuous Programming Part One: Linear Objectives," *J. Math. Anal. Appl.*, Vol. 28, pp. 32–51, 1969.

[39] R. C. Grinold, "Symmetric Duality for Continuous Linear Programs," *SIAM J. Appl. Math.*, Vol. 18, pp. 84–97, 1970.

[40] W. H. Hager and S. K. Mitter, "Lagrange Duality Theory for Convex Control Problems," *SIAM J. Control and Optimization*, Vol. 14, pp. 843–856.

[41] W. H. Hager, "Lipschitz Continuity for Constrained Processes," *SIAM J. Control and Optimization*, Vol. 17, pp. 321–338, 1979.

[42] B. Hajek and R. G. Ogier, "Optimal Dynamic Routing in Communication Networks with Continuous Traffic," *Networks*, Vol. 14, pp. 457–487, 1984.

[43] J. M. Harrison and J. G. Dai, "Solving Steady State Densities of Reflected Brownian Motion," *The Annals of Applied Probability*, Vol. 1, No. 1, pp. 16–35, 1991.

[44] J. M. Harrison, *Brownian Motion and Stochastic Flow Systems*, Graduate School of Business, Stanford University, Robert E. Krieger Publishing Company, Inc., Malabar, Florida, 1990.

[45] J. Hartberger, "Representation Extended to Continuous Time," *Optimization Methods for Resource Allocation*, R. W. Cottle and J. Krarup, eds., Crane Russak and Co. Inc., New York, pp. 309-322, 1974.

[46] A. E. Ilyutovich, "Iterative Optimization Method for Linear Programming Problems in Functional Spaces," *Studies in Linear and Nonlinear Programming*, Stanford, 1976.

[47] A. E. Ilyutovich, "Piecewise–Continuous Solutions of Linear Dynamic Problems in Economic Planning," *Auto. and Remote Control*, Vol. 41, pp. 501–508.

[48] S. Ito, C.T. Kelley and E.W. Sachs, "Inexact Primal–Dual Interior Point Iteration for Linear Program in Function Spaces," *Tech. Report*, Institute of Statistical Mathematics, Department of Prediction and Control, Japan, 1994.

[49] James R. Jackson, "Jobshop–like Queueing Systems," *Management Science*, Vol. 10, No. 1, October, 1963.

[50] Ram P. Kanwal, *Generalized Functions*, Academic Press Inc., 1983.

[51] F. P. Kelly, *Reversibility and Stochastic Processes*, Wiley, Chichester, 1979.

[52] F. P. Kelly, "Bounds on the Performance of Dynamic Routing Schemes for Highly Connected Networks," working paper 92–34, Statistics Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England, UK.

[53] L. Kleinrock, *Queueing Systems Volume* II: *Computer Applications*, John Wiley & Sons, 1976.

[54] G. P. Klimov, "Time Sharing Service Systems I," *Theory of Prob. and Its Applications*, Vol. 19, pp. 532–551, 1974.

[55] M. Kojima, T. Noma and A. Yoshise, *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Research Report RJ 7493, IBM Almaden Research Center, San Jose, CA, 1990.

[56] P. R. Kumar, "Re–Entrant Lines," preprint.

[57] R. S. Lehman, "On the Continuous Simplex Method," RM–1386, Rand Corporation, Santa Monica, CA, 1954.

[58] N. Levinson, "A Class of Continuous Linear Programming Problems," *J. Math. Anal. Appl.*, Vol. 16, pp. 73–83, 1966.

[59] Y. Y. Lin and J.-S. Pang, "Iterative Methods for Large Convex Quadratic Programs: A Survey," *SIAM J. Contr. & Optim.*, Vol. 25, pp. 383–411, 1987.

[60] D. G. Luenberger, *Optimization by Vector Space Method*, Wiley, New York, 1969.

[61] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison–Welsley, Reading, MA, 1973.

[62] Z.-Q. Luo and P. Tseng, "Error Bound and Convergence Analysis of Matrix Splitting Algorithms for the Affine Variational Inequality Problem," *SIAM Journal on Optimization*, Vol. 2, pp. 43–54, 1992.

[63] Thomas Magnanti, Jerry Shapiro and Mike Wagner, "Generalized Linear Programming Solves the Dual," *Management Science*, Vol. 22, No. 11, pp. 1195-1203, 1976.

[64] B. Martos, *Nonlinear Programming, Theory and Methods*, North–Holland Publishing Company, 1975.

[65] Franklin H. Moss and Andrian Segall, "An Optimal Control Approach to Dynamic Routing in Networks," *IEEE Trans. on Automatic Control*, Vol. AC–27, No. 2, April, 1982.

[66] Franklin H. Moss, "The Application of Optimal Control Theory to Dynamic Routing in Data Communication Networks," *Ph. D. thesis*, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 1977.

[67] K. G. Murty, *Linear Complementarity, Linear and Nonlinear Programming*, Heldermann, Verlag, Berlin, 1988.

[68] G. F. Newell, *Applications of Queueing Theory*, Chapman and Hall, 1982.

[69] C. H. Papadimitriou and J. N. Tsitsiklis, "The Complexity of Optimal Queueing Network Control," Extended Abstract, University of California at San Diego, La Jolla, CA, 1993.

[70] S. Papageorgiou, "A Class of Infinite Dimensional Linear Programming Problems," *J. Math. Anal. Appl.*, Vol. 87, pp. 228–245, 1982.

[71] J. R. Perkins and P. R. Kumar, "Optimal Control of Pull Manufacturing Systems," *tech. rep.*, Urbana, IL 61801, 1994.

[72] A. F. Perold, "Fundamentals of a Continuous Time Simplex Method," *Tech. Report Sol 78-26*, Department of Operations Research, Stanford University, Stanford, CA, 1978.

[73] A. F. Perold, "Extreme Points and Basic Feasible Solutions for Continuous Time Linear Programming," *SIAM J. Control and Opt.*, Vol. 19, pp. 52–63, 1981.

[74] A. B. Philpott and M. Craddock, "An Adaptive Discretization Algorithm for a Class of Continuous Network Programs," Department of Engineering Science, University of Auckland, New Zealand, submitted for publication, 1994.

[75] A. B. Philpott, "Network Programming in Continuous Time with Node Storage," In *Infinite Programming.* Proceedings of an International Symposium on Infinite Dimensional Linear Programming, E. J. Anderson & A. B. Philpott (Eds.), Springer–Verlag, New York, 1985.

[76] M. C. Pullan, A. Mason, M. Craddock, and A.B. Philpott, "Computer Programs for Solving the Capacitated Transshipment Problem," University of Auckland, School of Engineering Report 521, 1992.

[77] M. C. Pullan, "An Algorithm for a Class of Continuous Linear Programming Problems," *SIAM J. Control Optim.*, Vol. 31, pp. 1558–1577, 1993.

[78] M. C. Pullan, "Forms of Optimal Solutions for Separated Continuous Linear Programs," to appear in *SIAM J. Control Optim.*, 1994.

[79] M. C. Pullan, "A Duality Theory for Separated Continuous Linear Programs," to appear in *SIAM J. Control Optim.*, 1994.

[80] Michael Reed and Barry Simon, *Functional Analysis I, Methods of Modern Mathematical Physics*, Academic Press, Inc., 1980.

[81] R. T. Rockafellar, "State Constraints in Convex Control Problems of Bolza," *SIAM J. Control and Optimization*, Vol. 10, pp. 691–715, 1972.

[82] H. L. Royden, *Real Analysis*, (2nd ed.), Macmillan, New York, 1968.

[83] W. Rudin, *Real and Complex Analysis*, McGraw–Hill, New York, 1966.

[84] R. G. Segers, "A Generalized Function Setting for Dynamic Optimal Control Problems," *Optimization Methods for Resource Allocation*, R. W. Cottle and J. Krarup, eds., Crane Russak and Co. Inc., New York, pp. 279–296, 1974.

[85] E. Seneta, *Non–Negative Matrices and Markov Chains*, 2nd ed. New York: Springer–Verlag, 1981.

[86] S. P. Sethi, W. P. Drews and R. G. Segers, "A Unified Framework for Linear Control Problems with State–Variable Inequality Constraints," *J. Optim. Theory and Applications*, Vol. 36, pp. 93–109, 1982.

[87] Suresh P. Sethi, Qing Zhang, and Xun Yu Zhou, "Hierarchical Production Controls in a Stochastic Two–Machine Flow Shop with a Finite Internal Buffer," *tech. report*, University of Toronto, Ontario, Canada, July 1993.

[88] S. P. Sethi and X. Y. Zhou, "Stochastic Dynamic Job Shops and Hierarchical Production Planning," working paper, University of Toronto, 1992.

[89] Ali Sharifnia, "Stability and Performance of Distributed Control Methods Based on Continuous–Flow Models," *IEEE trans. on Automatic Control*, Vol. 39. No. 4, pp. 725–737, April, 1994.

[90] Ali Sharifnia, "Starvation–Based Instability of Scheduling Policies in Non–Acyclic Fluid and Queueing Networks," to appear in *IMA workshop on DEDs, Manuf. Systems and Communication Networks.*

[91] Ali Sharifnia, "Stability and Performance of a Simple Distributed Tracking Policy for Production Control of Manufacturing Systems," submitted to *IEEE trans. on Automatic Control.*

[92] W. F. Tyndall, "A Duality Theorem for a Class of Continuous Linear Programming Problems," *SIAM J. Appl. Math.*, Vol. 13, pp. 644–666, 1965.

[93] W. F. Tyndall, "An Extended Duality Theorem for Continuous Linear Programming Problems," *SIAM J. Appl. Math.*, Vol. 15, pp. 1294–1298, 1967.

[94] R. J. Vanderbei, *LOQO User's Manual*, Program in Statistics & Operations Research, Princeton University, Princeton, N.J. 08544.

[95] J. Walrand, *An Introduction to Queueing Networks*, Printice Hall, Englewood Cliffs, N.J., 1988.

[96] Lawrence M. Wein, "Scheduling a Two–Station Multi–Class Queueing Network in Heavy Traffic," *Sloan School of Management Working Paper #2016–88*, Massachusetts Institute of Technology, Cambridge, MA, 02139, November, 1987.

[97] Lawrence M. Wein, "Dynamic Scheduling of a Multi–Class Make–To–Stock Queue," *Operations Research*, Vol. 40, No. 4, July–August, 1992.

[98] Gideon Weiss, "On the Optimal Draining of Re–Entrant Fluid Lines," *tech. rep.*, School of ISyE, Georgia Tech, Atlanta, GA, 1994.

[99] M. Yamasaki, "Duality Theorems in Mathematical Programmings and their Applications," *J. Sci. Hiroshima Univ.*, Ser. A-1. Vol. 32, pp. 331-356, 1969.

[100] Yinyu Ye, "A Fully Polynomial-Time Approximation Algorithm for Computing a Stationary Point of the General Linear Complementarity Problem," *Mathematics of Operations Research*, Vol. 18, No. 2, May 1993.