On-line Traffic Signalization using Robust Feedback Control

Tungsheng Yu

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Dr. Joseph A. Ball, Chair
Dr. Pushkin Kachroo
Dr. David L. Russell
Dr. James E. Thomson
Dr. Robert L. Wheeler

December 18, 1997
Blacksburg, Virginia

Keywords: H_∞ control, dissipative system, traffic signalization, intersection, hybrid system, Hamilton-Jacobi inequality, finite state machines, queueing network

On-line Traffic Signalization using Robust Feedback Control

Tungsheng Yu

(ABSTRACT)

The traffic signal affects the life of virtually everyone every day. The effectiveness of signal systems can reduce the incidence of delays, stops, fuel consumption, emission of pollutants, and accidents. The problems related to rapid growth in traffic congestion call for more effective traffic signalization using robust feedback control methodology.

Online traffic-responsive signalization is based on real-time traffic conditions and selects cycle, split, phase, and offset for the intersection according to detector data. A robust traffic feedback control begins with assembling traffic demands, traffic facility supply, and feedback control law for the existing traffic operating environment. This information serves the input to the traffic control process which in turn provides an output in terms of the desired performance under varying conditions.

Traffic signalization belongs to a class of hybrid systems since the differential equations model the continuous behavior of the traffic flow dynamics and finite-state machines model the discrete state changes of the controller. A complicating aspect, due to the state-space constraint that queue lengths are necessarily nonnegative, is that the continuous-time system dynamics is actually the projection of a smooth system of ordinary differential equations. This also leads to discontinuities in the boundary dynamics of a sort common in queueing problems.

The project is concerned with the design of a feedback controller to minimize accumulated queue lengths in the presence of unknown inflow disturbances at an isolated intersection and a traffic network with some signalized intersections. A dynamical system has finite L_2-gain if it is dissipative in some sense. Therefore, the H_∞-control problem turns to designing a controller such that the resulting closed loop system is dissipative, and correspondingly there exists a storage function.

The major contributions of this thesis include 1) to propose state space models for both isolated multi-phase intersections and a class of queueing networks; 2) to formulate H_∞ problems for the control systems with persistent disturbances; 3) to present the projection dynamics aspects of the problem to account for the constraints on the state variables; 4) formally to study this problem as a hybrid system; 5) to derive traffic-actuated feedback control laws for the multi-phase intersections.

Though we have mathematically presented a robust feedback solution for the traffic signalization, there still remains some distance before the physical implementation. A robust adaptive control is an interesting research area for the future traffic signalization.

ACKNOWLEDGEMENTS

I am sincerely grateful to my advisor, Dr. Joseph A. Ball, for his immense help, excellent guidance and encouragement throughout my research. He has helped me to learn, understand, and appreciate H_∞ control concepts and their applications. This thesis could not have been completed without him. I would like to thank Dr. Pushkin Kachroo for his support and generous consultation in traffic signalization. I am also indebted to the other members of my committee: David L. Russell, James E. Thomson, and Robert L. Wheeler for reading this thesis and offering their helpful insight. I wish to extend my sincere thanks to Dr. Martin V. Day for his interest and contribution in this project, whose great expertise in differential equations was invaluable in analyzing the solutions of the key Hamiltonian-Jacobi-Isaacs equations.

I would like to thank my supervisor, Robert L. Miller, at AT&T, who allowed me to take a week off to prepare the final presentation of this project.

I wish to thank my wonderful parents for their support and encouragement they have given me throughout my education. A very special thanks to my wife, Hong Shi, and two children, William Yu and Patrick Yu, who have displayed tremendous patience and sacrifice while I completed this work. Without their support this thesis would not have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The traffic signal affects the life of virtually everyone every day. People need traffic signals to ensure safety and mobility. The effectiveness of signal systems can reduce the incidence of delays, stops, fuel consumption, emission of pollutants, and accidents. The problems related to rapid growth in traffic congestion call for more effective traffic signalization using robust feedback control methodology.

Online traffic-responsive signalization is based on real-time traffic conditions and selects cycle, split, phase, and offset (see Table 2.2 for definitions) for the intersection according to detector data. A robust traffic feedback control begins with assembling traffic demands, traffic facility supply, and feedback control laws for the existing traffic operating environment. This information provides the input to the traffic control process which in turn provides an output in terms of the desired performance. This is a promising control concept for enhancing traffic signalization.

1.1 Problem Statement

The problem of traffic signal control has been studied extensively. A state of the art summary is giving in the Traffic Control Systems Handbook [FHWA 1996].

As traffic flows increase greater than the capacity of an intersection or arterial network, congestion is bound to spread, with queues growing in length and duration. Existing control methods function effectively when traffic demands are below saturation but deteriorate rapidly when severe congestion persists for a sustained period. Indeed the nature of traffic control for saturated conditions may be essentially different from that of normal control modes. This suggests that in order to maintain the optimal circulation of traffic in existing arterial networks there is a need to re-appraise certain efficient signal control strategies.

Under relatively free flow conditions the offset between the start of green for the principal route at successive intersections is determined by the average speed of traffic. In other words the signalization facilitates a smooth progression of vehicle platoons along arterial routes. However, as traffic grows, congestion increases and intersections become saturated. The initial queue at the start of each green begins to disrupt any forward progression and long queues from one intersection may eventually spill back and disrupt movements at upstream intersections. To minimize the impact of potentially disruptive queues, many people have made great effort in the area of signal timing optimization techniques, such as [Webster 1958], [Dunne and Potts 1964], [Gazis 1964], [Miller 65], [Longley 1968], [Pignataro etc. 1975], [Michalopoulos and Stephanopoulos 1978], [Gartner 1983], [Heydecker 1987] and others.

Analytically, only an isolated two-phase intersection and a simple arterial street with two intersections have been studied. These two problems were studied using Pontryagin's principle for obtaining optimal control assuming a predetermined arrival patterns of the incoming flows [Gazis 1964]. The problem for multi-phase and network level becomes highly complex. Mainly open loop techniques, rolling horizon techniques, and heuristic techniques have been used by researchers and engineers in the control design. In this thesis, we will deploy a new design approach, robust feedback control, in this area, which may provide an improved level of control by using better modeling of traffic variables and improved control algorithms.

As saturation is a situation where demand exceeds capacity it is necessary to know the demand distribution over the control period. The equipment to be used for saturated control will be different from the standard control devices which collect little information about incoming vehicles held in queues. The ultimate deployment of the new concept is dependent on the future incorporation of advanced information technology into traffic control and vehicle systems. In fact, the robust feedback control does require multipoint or even continuous detection around some intersections. In this research, we assume that the queue lengths of all links are observable at any point of time.

The optimization criterion for the control is to minimize the total number of vehicles on the controlled links of the network (or intersections) all the time. Generally this criterion is equivalent to minimize the total time spent by the vehicles in the controlled network including both waiting times in the queues and travel times between intersections. Based on this optimization criterion, we model the queue as the output element. We want to figure out how we can achieve the desired goal for the output of the plant by choosing the inputs under all possible conditions. The first step is to find a right mathematical model describing the behavior of the plant. The second step is to use mathematical tools to find suitable inputs for the plant based on measurements we make for outputs. The third step is to search for the control law for the model so that it leads to the robustness of the plant.

1.2 Objectives and organization of the thesis

The main objectives of this research include (1) understanding the basic concepts of traffic signal control systems;(2) appreciating nonlinear H_∞ control concepts; (3) presenting robust and demand-responsive control strategy for optimum signalization and queue management at the isolated intersection; and (4) deploying a new solution using H_∞ design approach for a class of queueing networks.

Chapter 1 presents the problem of traffic signalization and the objectives of this thesis. In next section, it reviews the basic control concepts of the signalized intersection and three generations of traffic signalization.

Chapter 2 describes the configuration of a traffic signal control system, analyzes the traffic saturation and some key traffic control parameters, then explains traffic controls at isolated intersections and at networked intersections. Finally, it explains two basic types of traffic signal control.

Chapter 3 summarizes the classical concepts of a nonlinear H_∞ control system. It describes the theory of dissipative systems and how to deal with persistent disturbances. It follows the L_2-gain theory and the available storage concept. Then it builds the relationship among these concepts (L_2-gain, available storage and dissipative systems), and also covers the theory of nonlinear state feedback H_∞ control. Finally, it presents Filippov's solutions and the hybrid system formulation.

Chapter 4 describes the details of the state space model of a two-phase intersection, proposes the projected dynamical system associated with the model, derives the minimal storage function and then discusses the Filippov solution of discontinuous vector fields. It then develops the feedback controller of the hybrid system.

Chapter 5 proposes state space models for multi-phase isolated intersections and formulates the H_∞ optimization problem for the intersection signal control, and presents the solution of the problem. It concludes with the general traffic-actuated control algorithm for n-phase intersections.

Chapter 6 develops the H_∞ optimization models for the traffic network with two signalized intersections, and then extends to a class of queueing networks.

Chapter 7 summarizes our work and gives some recommendations on future research and development. The robust adaptive control is an interesting research area in traffic signalization.

1.3 A brief review of traffic signal control

Signalized intersection control concepts include [FHWA 1996]:

- Isolated intersection control — Controls traffic without considering adjacent signalized intersections.
- Arterial intersection control (open network) — Provides progressive traffic flow along the arterial. In this case, the signals must operate as a system.
- Closed network control — Coordinates a group of adjacent signalized intersection, such as in the central business district of a city.
- Areawide system control — Treats all or a major portion of signals in a city as a total system. Isolated, open or closed network concepts may control individual signals within this area.

There are three generations of signal control, each more sophisticated than its predecessor. The three generations [McShane and Roess 1990] can be described as:

- the first generation of control.

  Signal timing patterns for the system are pre-determined by the use of off-line optimization programs based on historical traffic data. These patterns or control plans are stored in the program and during control, appropriate plans for each control area are selected on a time-of-day basis by direct operator selection, or by matching the control plan with the existing traffic condition as detected by the control system. The minimum duration of a control period is fifteen minutes, and plan changes are only permitted on the hour or quarter hour for convenient operation. An offset transition procedure is incorporated in the area control logic to provide smooth transition during control plan changes. However, this procedure is not based on any optimization concept and does not necessarily minimize the disturbance caused by a plan change in the system.

- the second generation of control.

  Signal parameters such as cycle, offset and split for optimal network control are computed on a real-time basis as a function of current and predicted traffic conditions. A well known example for this generation is the Traffic Adaptive Network Signal Timing Program (TANSTP)[Kessmann and Ganslaw 1973], which consists of the following components:

  1. A traffic prediction model to predict traffic volumes and speeds in the network for the ensuing fifteen-minute period.
  2. A subnetwork configuration mode to subdivide the network on-line into subnetworks based on cycle length and other timing requirements.
  3. An optimization routine to generate optimum signal timing patterns on-line for each subnetwork for each fifteen-minute control period by minimizing the certain network delay expression:

4. A critical intersection control routine to fine-tune signal timings at critical intersections by adjusting split and offset on a cycle-by-cycle basis.
5. An offset transition model to determine transition parameters to effect a smooth and rapid offset transition during a timing pattern change.
6. A boundary model to compute timing patterns for optimum subnetwork interfacing.

- the third generation control.

  The strategy was conceived as a highly responsive control with a much shorter control period than second generation and without the restriction of a cycle-based system. Third generation included a queue management control at critical intersections. The control should properly handle three different traffic regimes, namely light, moderate and congested conditions, that is, the system will dynamically respond to the changing traffic conditions. In the light flow control policy, a base cycle length will be used and offset and split will be optimized jointly and implemented on a long-term basis, say a five-to-ten-minute control period. The moderate flow policy will involve the use of a variable cycle length for a control period of three-to-five-minute in duration without the need for offset transition and network structuring. The control policy for congested flow will also use a variable cycle length, but it will provide more rapid response to traffic and involve dynamic network structuring. The cycle length at the critical intersection will be expanded as high as possible to maximize capacity. Special algorithms will be provided to ensure that every second of green time is fully utilized at the critical intersection, to alleviate spillback on all links along the congested route and avoid network locking, and also to prevent the shifting of congestion downstream.

The first generation control software has been implemented all over the world, and continuously enhanced to improve the control methods. There also exist some second generation control systems such as, UTCS and OPAC systems developed in US, SCOOT system from England, PBIL system in Germany, and SCAT system of Australia. The third generation control is in progress.

These commonly used control strategies are variations of a traffic responsive split-variation routine for critical intersection control and table look-up procedure for area control based on pre-determined signal timings. They are not robust or responsive to the entire spectrum of traffic conditions. Therefore, we turn to the on-line feedback solutions for signalization on intersections. The models will fully utilize the sensored traffic parameters and the actuation signals. In fact, the feedback control method may be the best vehicle to achieve on-line, demand-responsive traffic signal timing, which are precisely based on the current traffic conditions. All adaptive traffic control methods are using some sort of feedback as well.

Chapter 2

Traffic Signal Control Systems

The general objectives of traffic signal control are to improve safety and to enhance mobility for the drivers. Mathematically they must first be reduced to quantitative objectives. Typically one would like to set up the traffic control problem as a system optimization problem, by defining an objective function and seeking an optimum for this function. This function will depend on certain control variables and controllable parameters of the system. The quality of the traffic signal control at the intersection can be measured by the five levels of service in Table 2.1. The ideal level is 5, it indicates the highest throughput for the intersection.

There are two types of traffic controls which can be used to influence the driver:

hard controls by signalization, which the driver must obey under penalty of law; and

soft controls by advisories, which provide the driver with real-time information on current traffic conditions for their independent decisions of route selection.

A generalized block diagram of the traffic signal control system is shown in Figure 2.1. Each

Table 2.1: The Level of Service of a Signalized Intersection

Level	*Representation*
1	a dense platoon arriving at the beginning of red.
2	a dense platoon arriving at the middle of red or a dispersed platoon arriving throughout the red.
3	random arrivals.
4	a dense platoon arriving in the middle of the green or a dispersed platoon arriving throughout the green.
5	a dense platoon arriving at the beginning of green.

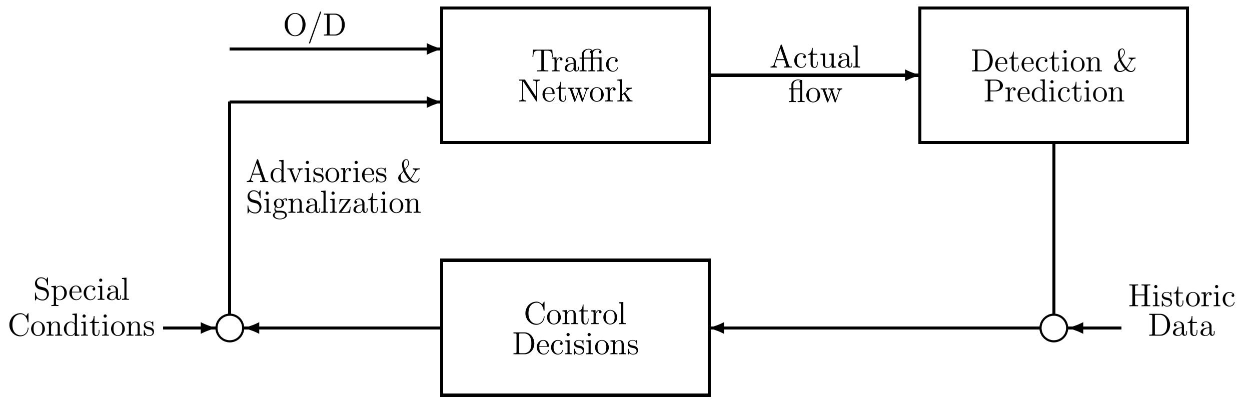

Figure 2.1: The Configuration of a Traffic Signal Control System

block represents a major function. The overall operation may be described as follows:

Traffic Network block. Based on their knowledge of origin/destination(O/D), the traffic condition and the signalization of the network, the drivers select their favorable routes, resulting in the *actual flows*. The process can be influenced by special conditions such as accidents and incidents which limit the available capacity in the network.

Detection/Prediction System block. Advanced information technology enables us to measure counts, speeds, occupancy, and queues to varying degrees of precision and accuracy, depending upon the number and placement of the point detectors in common use. The output of the block is the *detected flows*, which may not be true flow, due to the nature of the detectors used.

Control Decisions block. This block performs the major function of the control system – the decisions on control settings, *signalization* and *advisories*, by means of a "control policy". The control policy is a decision process on how to use the available information (historic data from traffic data warehouse, current data from the detection block, and special conditions from broadcasting) in order to make the optimal control settings.

2.1 Control variables and signal timing parameters

There are two types of control variables

Traffic control variables: describe traffic conditions, and

Environmental control variables: describe certain environmental conditions which may affect traffic performance.

Table 2.2: Signal Timing Variable Definitions

Variable	*Definition*
Interval	A discrete portion of the signal cycle during which the signal indications remain unchanged.
Cycle Length	The time required for 1 complete sequence of signal intervals (phases).
Phase	The portion of a signal cycle allocated to any single combination of one or more traffic movements simultaneously receiving the right-of-way during one or more intervals.
Split	The percentage of a cycle length allocated to each of the various phases in a signal cycle.
Offset	The time difference between the start of the green indication at one intersection as related to the system time reference point.

Traffic control variables include vehicle presence, flow rate, occupancy and density, speed, headway, and queue length. Environmental control variables include pavement surface (wet or icy), weather (rain, snow, or fog), and vehicle emissions. These control variables are generally measured from detector data. Table 2.2 summarizes the definitions of the fundamental signal timing variables [FHWA 1996], which are the key parameters for this research.

Cycle length should be constrained by the amount of storage available on the upstream link. The length of queue at the link should not be larger than the available link storage. That is,

$$qCd \leq l_i \tag{2.1}$$

where
q = flow in vehicle per second,
C = cycle length in seconds,
d = average vehicle storage length (meters per vehicle), and
l_i = available link storage (meters) at link i, that is, the maximum queue allowed at link i.

Let $l_{max} = \max_i \{l_i\}$, thus by equation (2.1) we have the maximum permissible cycle length

$$C_{max} = \frac{l_{max}}{qd}.$$

Remark 1: Cycle length may also be constrained by the amount of storage available on the downstream link. The cycle length must ensure that the flow being discharged from the current intersection will not cause congestion to occur at the downstream intersection.

Remark 2: Longer cycle lengths result in a small gain in capacity, but this small gain is offset by the disadvantages resulting from increasing delay and longer queues. It can be shown by Webster's method [Webster 1958] that to achieve even a modest increase of 3%

in vehicle flow, it would be necessary to increase the cycle length by 70% which is usually impossible to achieve in practice [OECD 1981].

2.2 Traffic control at isolated intersections

Vehicle total delay at the isolated intersection results from:

- Stopped time delay (time waiting during red), and
- start-up delay

The ideal control objectives at the intersection are:

- Maximize intersection capacity to lessen delay, and
- Reduce the potential for accident-producing conflicts.

However, these two objectives are generally incompatible [FHWA 1996]. Fewer phases and shorter cycle length lessens delay, while reducing accidents may require more phases and longer cycles. Therefore, it is necessary to apply sound engineering judgement to achieve the best possible compromise among these objectives.

To develop effective control concepts and strategies for isolated intersections, one must consider both traffic flow fluctuations and the random nature of vehicle arrivals. It has been shown that a Poisson distribution best predicts vehicle arrivals for isolated intersections [Drew 1963] in the normal condition. It gets a lot of disturbances in the oversaturated condition, however.

2.3 Traffic control at signal networks

There are the two basic types of signal networks

Open network: an arterial street signal system in which signals in a linear pattern are supervised to give progressive flow in one or two traveled directions. See Figure 2.2.

Closed network: a closed grid pattern signal system in which several arterials, forming an interlocking pattern, are supervised to give progressive flow in all traveled directions with the network to the extent practical. See Figure 2.3

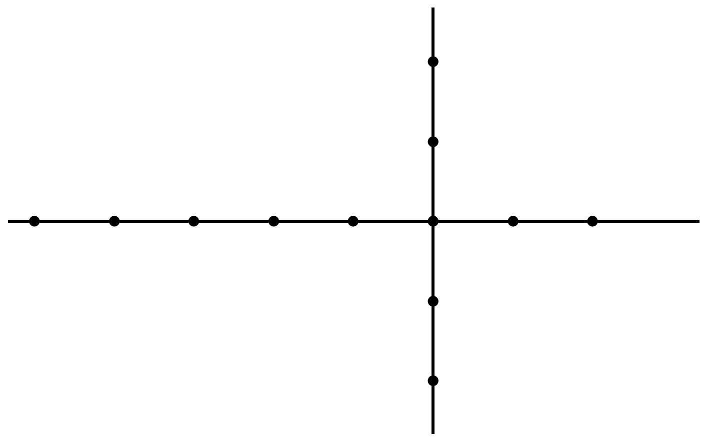

Figure 2.2: Open Signal Network

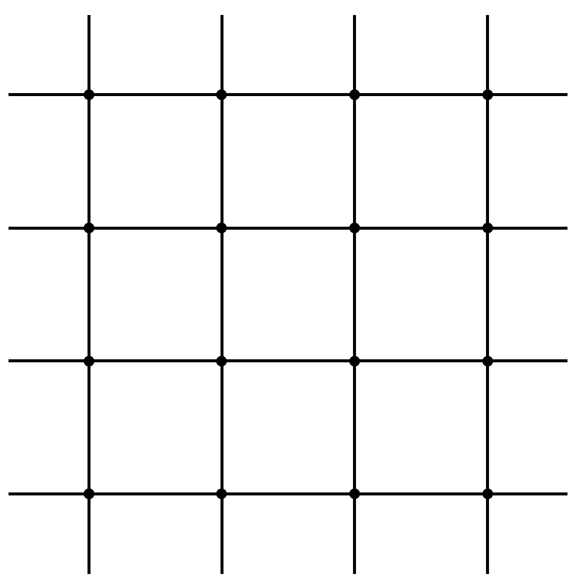

Figure 2.3: Closed Signal Network

The methodology of traffic control at signal networks has followed from the traditional methodologies of progression and traffic actuation. Thus, a typical traffic control system works on the following principle: One tries to synchronize as many traffic lights as possible in order to provide the opportunity of uninterrupted flow to as many cars as possible. The synchronization scheme is selected for average traffic conditions, and changes relatively infrequently, sometimes no more than three or four times during the day. In any case, the time scale of switching from one synchronization scheme to another is of the order of minutes or even hours. This type of control is referred to as *macrocontrol*. The problem associated with macrocontrol is to find a good synchronization scheme for given traffic conditions. On the other hand, one has to pay special attention to critical intersections where the time scale of signal switching is of the order of seconds. This is referred to as *microcontrol*. The problem associated with microcontrol is to find an algorithm that may adjust the traffic lights in response to variations of traffic demand, in order to minimize the delay in the neighborhood of the critical intersections and to limit the size of queue so that upstream intersections are not blocked.

Network signal control systems may adjust signal timing by:

- Selecting precomputed timing plans according to time-of-day,
- Selecting timing plans from the data base with alternatives using traffic detectors, and
- Generating real-time plans using traffic detectors.

2.4 Saturated Flow Conditions and Saturation

A *saturated flow condition* develops when demand at a point (or points) in a network exceeds capacity for a sustained period. This condition reveals itself at an intersection through the development of long queues, which may reach from one intersection to another.

Saturation is characterized by a discontinuity in the variation of the quality of service, or the abrupt drop in traffic characteristics. Saturation is identified by Table 2.3.

The traffic situation can be shown in terms of quality of service and demand as Figure 2.4 [OECD 1981]. As the demand increases, the quality of service decreases progressively from 1 to 2, where a queue builds up. It is at point 2 that the potential demand becomes greater than capacity. The fall 2-3 is due to two factors: (1) by the time for which the demand persists and (2) by the virtually instantaneous reduction in capacity. even if the demand decreases beyond 3, the system is not reversible and the return path towards a saturated situation lies below the outward path.

Since traffic flows along a system, made of some interdependent elements, saturation may appear either in an isolated intersection of the network without affecting the whole system

Table 2.3: Identification of Saturation

Perspective	*Indicators*
off-the-road environment	exhaust pollution and noise level amenity of the family dwellings in the area
pedestrians and cyclists	accident rate number of hourly impeded crossing opportunities pedestrian accident indices
in-vehicle travellers	low speeds and long travel times excessively long travel delays intersection delays in excess of one cycle queue length at intersections number of side street blockings number of stops per link high densities and high volume/capacity ratios

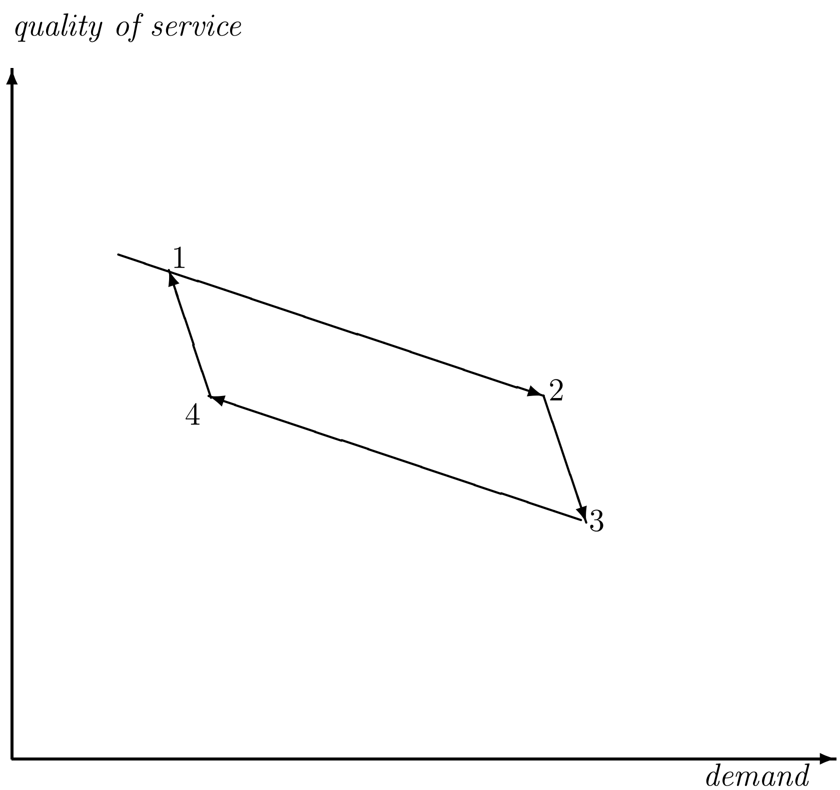

Figure 2.4: Traffic Saturation Point

Table 2.4: Traffic Condition Descriptions

Undersaturated Condition	*Saturated Condition*		*Jam Condition*
	Stable	*Unstable*	
No queue formation No delay	Queue formation but not growing. Delay effects are local	Queue formation and growing. Delay effects are are still local and duration is short	Queue formation and growing to point where upstream intersection performance adversely affected

(localized saturation) or, because of interdependent, may be transmitted to the rest of the system, substantially disrupting its functional capacities (jam), without every affected intersection necessarily reaching its congestion threshold. Traffic condition description is shown in Table 2.4 on the basis of the queue formation mechanism.

Saturation problems are usually local in nature at their onset. If demand continues at or above capacity, saturation will spread and symptoms will become evident throughout the remainder of the arterial or network, involving completely loaded links and spillback through and beyond upstream intersections. That is, a jam condition is developing.

There are two basic rules to solve the problems: increase capacity and reduce demand. However, it becomes more difficult to physically increase the capacity and reduce the demand. The reality is that we have to develop new control concepts such that the capacity may remain greater than demand. Armed with the feedback solution, one can detect potential problems before service levels decline rapidly, and make intelligent decisions about signal settings.

2.5 Types of traffic signal control

There are two basic types of traffic signal control for isolated intersections:

Pretimed control: assigns right-of-way according to a predetermined schedule, and

Traffic-actuated control: assigns right-of-way according to real-time measures of traffic demand obtained from vehicle detectors placed on one or more approaches. It adjusts green times and possibly phase sequence (i.e., skip-phase).

Each type offers varying performance and cost characteristics depending on the installation and prevailing traffic conditions. Some controllers for the networked intersections may use any combination of both controls based on the level of traffic flow and the need of signal coordination.

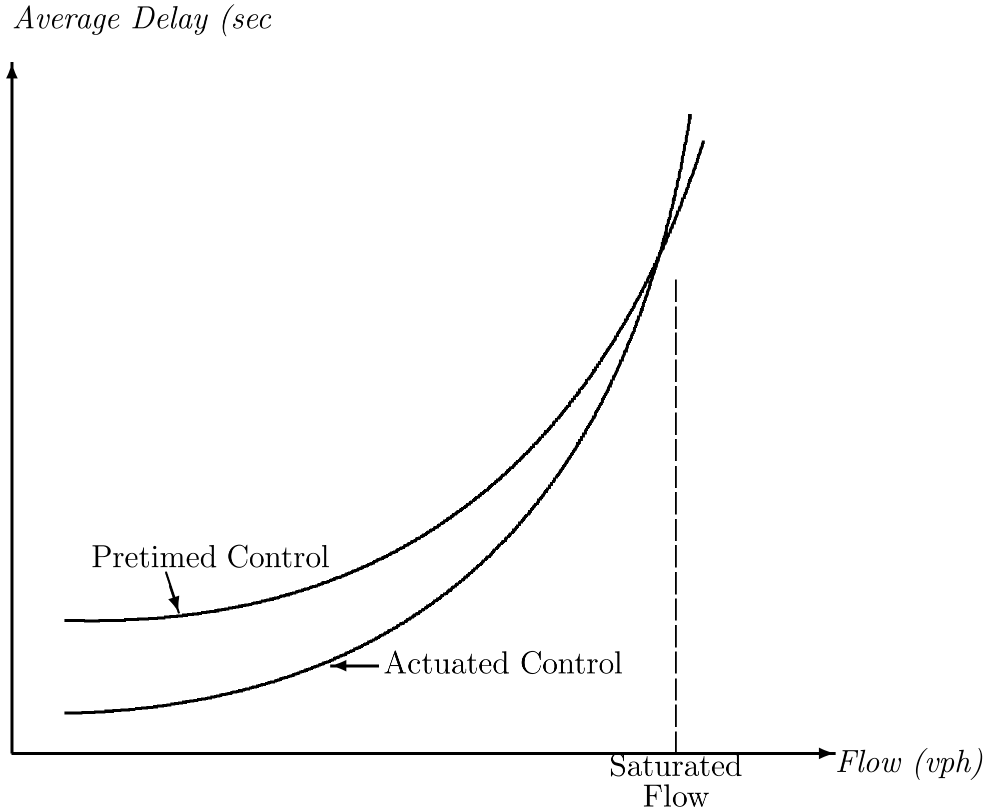

Figure 2.5: Delay-Flow Curves for Actuated Control vs. Pretimed Control

It has been shown, under moderate flows, traffic-actuated control generally proves more effective than pretimed control [McShane and Roess 1990]. However, the current traffic-actuated control is acting as the pretimed control under saturated conditions [OECD 1981]. Figure 2.5 presents the delay-flow curves between the traffic-actuated control and the pretimed control. When a jam condition occurs, traffic cannot move even when it receives a green light. To clear traffic during jam conditions requires a different concept of control.

Using traffic-actuated control at isolated intersections enables the timing plan to adjust continuously in response to traffic demand. A phase may be skipped when no demand is present at the associated direction. As long as the time between vehicle actuations remain shorter than the vehicle interval, green will be retained on that phase subject to the maximum green interval. However, the potential values to be realized depend on

1. Algorithm of control operation
2. Location of detectors,
3. Controller timing settings (i.e., minimum green time, passage time, and maximum green time), and

4. Type of equipment installed.

The scope of this thesis is to find proper algorithms of operation for traffic-actuated control. The other three issues can be found in detail in [FHWA 1996].

The literature on actuated signal operations is limited compared with that on fixed-time signals. The descriptions of actuated controller operations are provided by [Staunton 1976]. The analytical method for estimating average green times and cycle time at vehicle-actuated signals are provided by [Lin 1982] and [Akcelik 1994].

Chapter 3

H_∞ theory — Robust Control Systems

H_∞ theory addresses the *robust control* of achievable time-domain performance under uncertainty. Robust control means designing a feedback control system that delivers a desirable performance not only for the nominal process, but also for the process under all kinds of disturbances. It is our goal to deploy H_∞ design methodology to produce our feedback solutions.

A system is described as a process seen as a part of reality. Given five finite-dimensional Euclidean spaces: the *state space ($\mathcal{X}$)*, the *control space ($\mathcal{U}$)*, the *disturbance space ($\mathcal{D}$)*, the *measurement space ($\mathcal{Y}$)*, the *error space ($\mathcal{Z}$)*, and three continuous functions:
$f : \mathcal{X} \times \mathcal{U} \times \mathcal{D} \to \mathcal{X}$, $\quad g : \mathcal{X} \times \mathcal{D} \to \mathcal{Y}$, $\quad$ and $\quad h : \mathcal{X} \times \mathcal{U} \to \mathcal{Z}$,
the system (Σ) can be mathematically expressed in the form:

$$\Sigma : \begin{cases} \dot{x} &= f(x, u, d) \\ y &= g(x, d) \\ z &= h(x, u) \end{cases} \tag{3.1}$$

As shown in Figure 3.1, Σ represents a standard control system with two sets of inputs $u \in \mathcal{U}$ and $d \in \mathcal{D}$, two sets of outputs $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$, and state variable $x \in \mathcal{X}$. Associated with Σ are admissible measurements, admissible disturbances, and admissible controls, each characterized by a set-valued constraint.

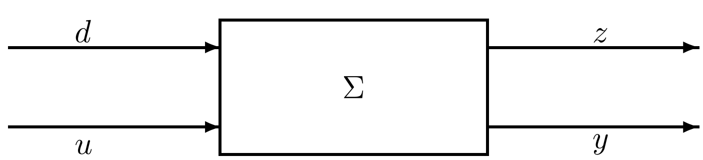

Figure 3.1: Standard Control Configuration

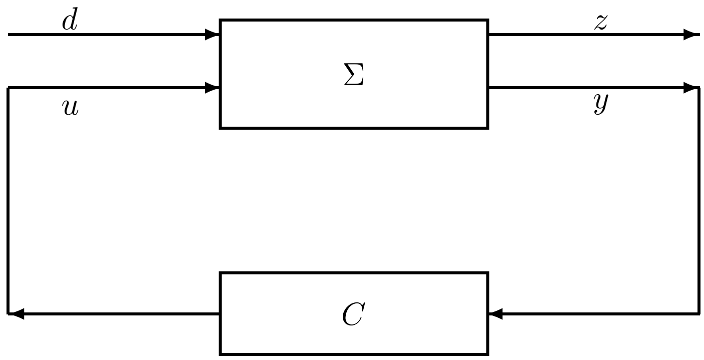

Figure 3.2: Standard Closed-loop Configuration

The system is subjected to *exogenous* inputs d such as disturbances-to-be-rejected or reference signals to-be-tracked. These inputs represent the input uncertainty. The system has a set of *control* inputs u and a set of *measured* outputs y, both accessible for feedback. The process is viewed as an element of a larger set of plants that reflects the modeling errors. The error signals that capture the performance specifications are called *to-be-controlled* outputs z. These outputs represent the quantity which enter explicitly into the formulation of the control objectives. Typical performance specifications involve good disturbance rejection, good tracking, limits on the control authority and bandwidth, etc. The *state* x of a system summarizes the information about the system and the external signals acting on the system insofar as relevant for the future behavior. It provides a formal link between inputs and outputs. We shall use the measurement y to choose our input u so that it will minimize the effect of d on z.

A strong constraint we impose here is the robustness of the system. *Robustness* of a system says nothing more than that the mapping from y to u should guarantee the internal stability of the closed loop system no matter which admissible disturbance is present. In other words, robustness leads to guaranteed performance even if the worst disturbance in the prescribed parameter range happens.

The *controller* C is a decision maker which advises the steering authority of a real control system on how to specialize the control variable at each time. The decision should be based exclusively on past measurements of the output and should take inaccuracies in the measurement into account. A decision procedure which meets this requirement is referred to as *feedback*. In other words, a feedback can be viewed as a mechanism that corrects the inputs of the process based on the known results of the outputs.

There are two classes of feedback, namely *state feedback* and *measurement feedback*. State feedback can be used when the state can be measured completely (i.e. $y = x$); while measurement feedback will be used when the measurement is a possible disturbed function of the state.

The *optimal H_∞ control problem* is to find a Controller C, processing the measurements y and producing the control inputs u, such that in the closed-loop configuration of Figure 3.2 the L_2-gain from exogenous inputs d to to-be-controlled outputs z is minimized, and furthermore, the close-loop system is stable "in some sense".

The optimal H_∞ control problem is usually hard, so people turn to the *suboptimal H_∞ control problem*, which is to find, if possible, for a given disturbance attenuation level γ a controller C such that the closed-loop system has L_2-gain $\leq \gamma$, and is stable. The solution to the optimal H_∞ control problem may then be approximated by an iteration of the suboptimal H_∞ control problem through successively decreasing γ to the optimal disturbance attenuation level.

The terminology H_∞ [Zames 1981] stems from the fact that in the linear case the L_2-gain of a stable system is equal to the H_∞ norm of its transfer matrix. That is, the H_∞ norm from the exogenous disturbance inputs to the to-be-controlled variables in the frequency domain, used to describe the control objectives, is equal to the L_2-induced operator norm for the time-domain versions, under the constraint of internal stability.

The property of finite L_2-induced norm of a stable system can be characterized as the dissipativity [Willems 1972] of the system with respect to a certain supply rate. A dynamical system has finite L_2-gain if it is dissipative in an appropriate sense to be described below. Therefore, the H_∞-control problem turns to designing a controller such that the resulting closed loop system is dissipative.

3.1 Dissipative systems with persistent disturbances

From the standpoint of the robust control, it is a natural objective to require a physical system which dissipates energy. Consider a nonlinear system of the following state space form

$$\Sigma : \begin{cases} \dot{x} = f(x,d), & x \in \mathbf{R}^n,\, d \in \mathbf{R^m} \\ z = h(x), & z \in \mathbf{R^p} \end{cases} \tag{3.2}$$

where f and h are at least continuously differentiable. Without loss of generality we assume that the system (3.2) has an equilibrium in $(x,d) = (0,0)$, i.e. $f(0,0) = 0$, and $h(0) = 0$.

We denote the *supply rate* as a locally integrable function

$$s : \mathcal{D} \times \mathcal{Z} \to \mathbf{R}^+$$

where $\mathcal{D} \subset \mathbf{R^m}$ and $\mathcal{Z} \subset \mathbf{R^p}$.

Let $\mathcal{X}$ be a connected subset of $\mathbf{R^n}$ containing the origin.

Definition 3.1 *A state space system Σ in (3.2) is said to be* **dissipative** *with respect to the supply rate s if there exists a function $S : \mathcal{X} \to \mathbf{R}^+$ such that for all $x_0 \in \mathcal{X}$, all $t_1 > t_0$, and* *all $d(\cdot)$, we have*

$$S(x(t_1)) \leq S(x(t_0)) + \int_{t_0}^{t_1} s(d(t), z(t))dt \tag{3.3}$$

where $x(t_0) = x_0$ is the initial state, and $x(t_1)$ is the state of Σ at time t_1 resulting from x_0 and $d(\cdot)$. We call this $S(\cdot)$ the **storage function** *and the inequality (3.3) as the* **integral dissipation inequality**.

Remark 1: The inequality (3.3) shows that the stored energy $S(x(t_1))$ of Σ at any future time t_1 will never be more than the sum of the initial energy $S(x_0)$ and the *total* externally supplied energy $\int_{t_0}^{t_1} s(d(t), z(t))dt$ during the period. In other words, Σ cannot generate any internal energy but possibly dissipate internal energy.

Remark 2: For some systems, the total externally supplied energy will become huge as the length of the time interval goes to infinity. It may make sense for us to use the *average* externally supplied energy instead of the total externally supplied energy in the inequality (3.3). So our argument is that *the energy change per unit time from initial state x_0 to the any future state x_1 will not exceed the externally supplied energy per unit time during the period.* That is,

$$\frac{1}{t_1 - t_0} \left(S(x(t_1)) - S(x(t_0)) \right) \leq \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} s(d(t), z(t))dt \tag{3.4}$$

We choose the supply rate as

$$s(d(t), z(t)) = \frac{1}{2} \left(\gamma^2 \|d(t)\|^2 - \|z(t)\|^2 \right), \qquad \gamma \geq 0 \tag{3.5}$$

and assuming $x(0) = 0$ and $S(0) = 0$. Then Σ is dissipative with respect to (3.5) if and only if there exists $S(\cdot) \geq 0$ such that for all $t_1 \geq t_0$, $x(t_0)$ and $d(\cdot)$,

$$\frac{1}{2} \int_{t_0}^{t_1} \left(\gamma^2 \|d(t)\|^2 - \|z(t)\|^2 \right) dt \geq S(x(t_1)) - S(x(t_0)) = S(x(t_1)) \geq 0$$

and let $t_1 = T$, $t_0 = 0$ this becomes

$$\int_0^T \|z(t)\|^2 dt \leq \gamma^2 \int_0^T \|d(t)\|^2 dt$$

Two sides divided by $T > 0$, we have

$$\frac{1}{T} \int_0^T \|z(t)\|^2 dt \leq \frac{\gamma^2}{T} \int_0^T \|d(t)\|^2 dt \tag{3.6}$$

Physically there are a lot of signals such as traffic arrival flows into an intersection which are periodic and bounded but not decaying at infinity in time. For these systems it is often reasonable to assume that

$$\sup_T \frac{1}{T} \int_0^T \|d(t)\|^2 dt < \infty$$

whereas it is unrealistic to assume that

$$\sup_T \int_0^T \|d(t)\|^2 dt < \infty$$

since the disturbance inputs $d(t)$ are persistent and not decaying as time goes to infinity.

3.2 L_2-gain, available storage and the properties of dissipative systems

We define the L_2-gain of the nonlinear system Σ in (3.2) and the available storage in the following way ([Willems 1972], [van der Schaft 1996]).

Definition 3.2 *Given a constant $\gamma \geq 0$, Σ is said to have L_2-**gain** $\leq \gamma$ if for all $x \in \mathbf{R^n}$ there exists a constant $\alpha(x)$, $0 \leq \alpha(x) < \infty$, with $\alpha(0) = 0$, such that the following inequality holds*

$$\int_0^t \|z(\tau)\|^2 d\tau \leq \gamma^2 \int_0^t \|d(\tau)\|^2 d\tau + \alpha(x_0) \tag{3.7}$$

for all $d \in L_2[0,t]$ and all $t \in [0,T]$, with $(0,T)$ any open interval in which the corresponding solutions $\phi(\tau, 0, x_0, d)$ of the differential equation $\dot{x} = f(x,d)$ exist, with $z(\tau) = h(\phi(\tau, 0, x_0, d))$ denoting the output of Σ resulting from $d(\cdot)$ and the initial state $x(0) = x_0$.

The L_2-gain of Σ is defined as

$$\gamma(\Sigma) = inf\,\{\gamma : \Sigma \ \text{ has } \ L_2\text{-}gain \leq \gamma\}$$

Σ is said to have L_2-gain$< \gamma$ if there exists $\gamma_0 < \gamma$ such that Σ has L_2-gain $\leq \gamma_0$.

Remark: The smallest such γ is also referred to as the L_2-induced norm from the disturbance inputs to the to-be-controlled outputs, which is mentioned in the beginning of this chapter. It is the time-domain interpretation of the nonlinear H_∞ norm.

Definition 3.3 *The* **available storage** *is defined as*

$$S^a(x) = \sup_d \frac{1}{2} \int_0^t (\|z(\tau)\|^2 - \gamma^2 \|d(\tau)\|^2) d\tau \tag{3.8}$$

where the supremum *is taken over all $d \in L_2[0,t]$ and all $t \geq 0$, and where $z(\tau)$ denotes the response of Σ to a disturbance d and an initial condition $x(0) = x$.*

Remark: The quantity $S^a(x)$ can be interpreted as the maximum amount of energy which can be extracted from the system Σ over all state trajectories connecting from the initial condition $x(0) = x$ to the resulting state $x(t) = x^*$.

From these definitions we easily obtain

Lemma 3.1 *Σ has L_2-gain$\leq \gamma$ if and only if the available storage S^a in (3.8) is finite for every $x \in \mathbf{R^n}$ and $S^a(0) = 0$.*

Remark: $2S^a(x)$ is the minimal constant $\alpha(x)$ for which the inequality (3.7) holds.

Lemma 3.2 *If S^a is finite for all $x \in \mathbf{R^n}$ then S^a is a minimal storage function. That is, any other possible storage function S satisfies $S(x) \geq S^a(x) \geq 0$ for all $x \in \mathbf{R^n}$*

Proof Suppose S^a is finite. Take $t = 0$ in (3.8) we see

$$S^a(x) \geq 0$$

for all $x \in \mathbf{R^n}$.

Compare now $S^a(x(t_0))$ with $S^a(x(t_1)) + \frac{1}{2}\int_{t_0}^{t_1}(\|z\|^2 - \gamma^2\|d\|^2)\,dt$, for a given $d \in L_2[t_0, t_1]$ taking $x(t_0)$ to $x(t_1)$ and then taking Σ from $x(t_1)$ at $t = t_1$ to state $x(t)$ at time t over all $d \in L_2[t_1, t]$. Since $S^a(x_0)$ is given as the supremum over all $d \in L_2[t_0, t_1]$ from the initial state x_0 to state $x(t)$ at time t, it follows that

$$S^a(x(t_0)) \geq S^a(x(t_1)) + \frac{1}{2}\int_{t_0}^{t_1}\left(\|z(t)\|^2 - \gamma^2\|d(t)\|^2\right)dt$$

that is,

$$S^a(x(t_1)) \leq S^a(x(t_0)) + \frac{1}{2}\int_{t_0}^{t_1}\left(\gamma^2\|d(t)\|^2 - \|z(t)\|^2\right)dt$$

and thus S^a is a storage function by Definition 3.1.

Suppose $S \geq 0$ is another storage function. Let $x(0) = x$, then

$$S(x(0)) + \frac{1}{2}\int_0^t\left(\gamma^2\|d(\tau)\|^2 - \|z(\tau)\|^2\right)d\tau \geq S(x(t)) \geq 0$$

that is,

$$S(x(0)) \geq \frac{1}{2}\int_0^t\left(\|z(\tau)\|^2 - \gamma^2\|d(\tau)\|^2\right)d\tau$$

The left side of the above inequality is independent to $d(\cdot)$. Therefore, we obtain

$$S(x(0)) \geq \sup_d \frac{1}{2}\int_0^t\left(\|z(\tau)\|^2 - \gamma^2\|d(\tau)\|^2\right)d\tau$$

By $x(0) = x$ and the definition of $S^a(x)$, this leads to

$$S(x) \geq S^a(x)$$

This completes the proof of the lemma.

By Lemma 3.1 and Definition 3.1, we easily obtain

Lemma 3.3 *Σ has L_2-gain$\leq \gamma$ if and only if Σ is dissipative with respect to the supply rate $s(d,z) = \frac{1}{2}\left(\gamma^2\|d\|^2 - \|z\|^2\right)$.*

In other words, Σ has L_2-gain$\leq \gamma$ if and only if there exists a nonnegative solution S to the integral dissipation inequality

$$S(x(t_1)) - S(x(t_0)) \leq \frac{1}{2}\int_{t_0}^{t_1}(\gamma^2\|d(\tau)\|^2 - \|z(\tau)\|^2)d\tau, \quad S(0) = 0 \tag{3.9}$$

for all $t_1 \geq t_0$, all $d \in L_2[t_0, t_1]$ and all $x \in \mathbf{R^n}$.

Now we assume that $S(\cdot)$ in (3.9) is continuously differentiable, then it immediately follows that this $S(\cdot)$ is a solution to the *differential dissipation inequality*

$$S_x(x)f(x,d) \leq \frac{1}{2}\gamma^2\|d\|^2 - \frac{1}{2}\|z\|^2, \quad S(0) = 0 \tag{3.10}$$

for all $d \in \mathcal{D}$ and all $x \in \mathbf{R^n}$.

Consider the function

$$S_x(x)f(x,d) - \frac{1}{2}\gamma^2\|d\|^2 + \frac{1}{2}\|z\|^2.$$

Its Hessian at $(x,d) = (0,0)$ with respect to d is equal to $-\gamma^2 I$. Therefore, in a neighborhood of the origin we can find the worst case disturbance with respect to the inequality (3.10) given by

$$d_{max}(x) = \arg\max_{d}\left(S_x(x)f(x,d) - \frac{1}{2}\gamma^2\|d\|^2 + \frac{1}{2}h^T(x)h(x)\right) \tag{3.11}$$

Thus locally the differential dissipation inequality (3.10) is equivalent to the *Hamilton-Jacobi inequality*

$$S_x(x)f(x,d_{max}(x)) - \frac{1}{2}\gamma^2\|d_{max}(x)\|^2 + \frac{1}{2}h^T(x)h(x) \leq 0, \quad S(0) = 0 \tag{3.12}$$

Note: The notion of Hamilton-Jacobi inequality will be formally defined in next section.

The properties of dissipative systems are summarized in the following theorem.

Theorem 3.1 (The properties of dissipative systems)

1. *The system Σ in (3.2) has L_2-gain$\leq \gamma$ if and only if there exists a solution $S : \mathbf{R^n} \rightarrow \mathbf{R}^+$ to the integral dissipation inequality (3.9) for all $t_1 \geq t_0$, all $d \in L_2[t_0, t_1]$ and all $x \in \mathbf{R^n}$.*

2. *There is a nonnegative C^1-solution to the integral dissipation inequality (3.9) if and only if there exists a nonnegative C^1-solution to the differential dissipation inequality (3.10) for all $d \in \mathcal{D}$.*

3. *There exists a local nonnegative C^1-solution to the differential dissipation inequality (3.10) if and only if locally there exists a nonnegative C^1-solution to the Hamilton-Jacobi inequality (3.12).*

Remark: If Σ is reachable from x^*, then we may define *the required supply from x^* as*

$$S^r(x) = \inf_d \frac{1}{2}\int_{-t}^{0}(\gamma^2\|d(\tau)\|^2 - \|z(\tau)\|^2)d\tau \tag{3.13}$$

where the infimum *is taken over all $d \in L_2[-t, 0]$ and all $t \geq 0$, with $x(-t) = x^*$ and $x(0) = x$.* The quantity $S^r(x)$ can be interpreted as the minimum amount of energy which is required in moving from the "ground state" $x(t) = x^*$ to any reachable point x in the state space over the interval $[-t, 0]$. We can prove that for all possible storage functions S, the bound $S^a(x) \leq S(x) \leq S^r(x) + S(x^*)$ holds [van der Schaft 1996]. This chain of inequalities has the interesting physical interpretation that *a dissipative system can supply to the outside only a fraction of what it has stored and can store only a fraction of what has been supplied to it.*

3.3 State feedback H_∞ control

The main topic of this thesis will be the state feedback H_∞ control problem. In this problem, it is assumed that the whole state is available for measurements (i.e. $g(x, d) = x$ in (3.1)) and the goal is to construct a static state feedback such that the closed loop system (3.1) has L_2-gain $\leq \gamma$ from d to z. To consider internal stability separately, we introduce the following two problems.

Definition 3.4 (state feedback L_2-gain optimal control problem): *Find, if possible, the smallest value γ^* such that for any $\gamma > \gamma^*$ there exists a state feedback*

$$u = l(x), \qquad l(0) = 0 \tag{3.14}$$

such that the L_2-gain of the closed loop system (3.1), (3.14) has L_2-gain $\leq \gamma$ (from d to z).

Definition 3.5 (state feedback H_∞ optimal control problem): *Find, if possible, the smallest value γ^* such that for any $\gamma > \gamma^*$ there exists a state feedback*

$$u = l(x), \qquad l(0) = 0 \tag{3.15}$$

such that the L_2-gain of the closed loop system (3.1), (3.15) has L_2-gain $\leq \gamma$ (from d to z), and the equilibrium is locally asymptotically stable.

The L_2-gain optimal control problem can be viewed as a two player, zero sum differential game, where one player u is called the minimizing player whose goal is to minimize the cost criterion

$$\frac{1}{2}\int_0^t (\|h(x(\tau), u(\tau))\|^2 - \gamma^2\|d(\tau)\|^2)d\tau$$

for every t; while the other player d is called the maximizing player whose goal is to maximize the same cost criterion.

The pre-Hamiltonian function associated with this game for the system (3.1) is a function $K_\gamma : T^*\mathcal{X} \times \mathcal{D} \times \mathcal{U} \to \mathbf{R}$ defined as

$$K_\gamma(x, p, d, u) = p^T f(x, u, d) - \frac{1}{2}\gamma^2\|d\|^2 + \frac{1}{2}\|h(x, u)\|^2 \tag{3.16}$$

where the cotangent manifold $T^*\mathcal{X}$ (at least locally) can be identified with $\mathcal{X} \times \mathcal{X}$, and $(x, p) \in T^*\mathcal{X}$.

Then we compute (if possible) a saddle point with respect to u and d in a neighborhood of the origin, i.e., there exist unique functions $d = d^*(x, p)$ and $u = u^*(x, p))$ defined around $(x, p) = (0, 0)$, satisfying

$$\frac{\partial K_\gamma}{\partial d}(x, p, d^*(x, p), u^*(x, p)) = 0$$
$$\frac{\partial K_\gamma}{\partial u}(x, p, d^*(x, p), u^*(x, p)) = 0$$

with $d^*(0, 0) = 0$ and $u^*(0, 0) = 0$, and satisfying the saddle point condition

$$K_\gamma(x, p, d, u^*) \leq K_\gamma(x, p, d^*, u^*) \leq K_\gamma(x, p, d^*, u) \tag{3.17}$$

for all state vectors x, costate vectors p, disturbances q and controls u around the origin.

The existence of the saddle point depends on regularity of the Hessian of K_γ with respect to d and u at the origin. This Hessian equals

$$\begin{pmatrix} -\gamma^2 I & 0 \\ 0 & (\frac{\partial h}{\partial u}(0,0))^T(\frac{\partial h}{\partial u}(0,0)) \end{pmatrix} \tag{3.18}$$

Thus if the derivative of h with respect to u has full column rank at the origin, local existence of the functions $d^*(x, p)$ and $u^*(x, p)$ follows directly from the Implicit Function Theorem.

Substituting the saddle point (d^*, u^*) into the pre-Hamiltonian K_γ leads to the Hamiltonian $H_\gamma : T^*\mathcal{X} \to \mathbf{R}$ defined as

$$H_\gamma(x, p) = K_\gamma(x, p, d^*(x, p), u^*(x, p))$$

The inequality

$$K_\gamma(x, S_x^T(x), d^*(x, S_x^T(x)), u^*(x, S_x^T(x))) \leq 0 \tag{3.19}$$

is well known as the *Hamilton-Jacobi inequality.*

The equation

$$H_\gamma(x, S_x^T(x)) = 0$$

is known as the *Hamilton-Jacobi-Isaacs equation*, sometimes, simply called the *Hamilton-Jacobi* equation.

We have the following well-known theorem.

Theorem 3.2 *Consider the system (3.1) with $y = x$ and $u^*(x,p), d^*(x,p)$ satisfying (3.17). Let $\gamma > 0$. Suppose there exists a local C^r $(k \geq r > 1)$ solution $S \geq 0$ to the Hamilton-Jacobi inequality given by (3.19). Then the C^{r-1} state feedback*

$$u = u^*(x, S_x^T(x)) \tag{3.20}$$

is such that the closed-loop system (3.1), (3.20), i.e.,

$$\begin{cases} \dot{x} &= f(x, u^*(x, S_x^T(x)), d) \\ z &= h(x, u^*(x, S_x^T(x))) \end{cases} \tag{3.21}$$

has L_2-gain$\leq \gamma$.

Conversely there exists a C^{r-1} feedback

$$u = l(x) \tag{3.22}$$

such that there exists a C^1 storage function $S \geq 0$ for the closed-loop system (3.1), (3.22) with supply rate $\frac{1}{2}\gamma^2\|d\|^2 - \frac{1}{2}\|z\|^2$. Then $S \geq 0$ is also a solution of the Hamilton-Jacobi inequality (3.19).

Proof Let $S \geq 0$ satisfy (3.19). By K_γ definition in (3.16) we substitute $p = S_x^T(x)$ in the first inequality of (3.17) to obtain

$$\begin{aligned} S_x(x) f(x, u^*(x, S_x^T(x)), d) - \frac{1}{2}\gamma^2\|d\|^2 + \frac{1}{2}\|h(x, u^*(x, S_x^T(x)))\|^2 \\ \leq K_\gamma(x, S_x^T(x), d^*(x, S_x^T(x)), u^*(x, S_x^T(x))) \end{aligned}$$

and thus by (3.19) for all $d \in \mathcal{D}$

$$S_x(x) f(x, u^*(x, S_x^T(x)), d) \leq \frac{1}{2}\gamma^2\|d\|^2 - \frac{1}{2}\|h(x, u^*(x, S_x^T(x)))\|^2$$

showing that (3.21) has L_2-gain$\leq \gamma$ with storage function S according to Theorem 3.1.

Conversely, let $S \geq 0$ be a C^1 storage function for the closed-loop system. By Theorem 3.1 we know that S is a solution to

$$S_x(x) f(x, l(x), d) \leq \frac{1}{2}\gamma^2\|d\|^2 - \frac{1}{2}\|h(x, l(x))\|^2 \tag{3.23}$$

for all $d \in \mathcal{D}$.

Then by substituting $p = S_x^T(x)$ and $u = l(x)$ in the second inequality of (3.17) we have

$$K_\gamma(x, S_x^T(x), d^*(x, S_x^T(x)), u^*(x, S_x^T(x))) \le K_\gamma(x, S_x^T(x), d^*(x, S_x^T(x)), l(x))$$

By (3.16), this becomes

$$\begin{aligned} K_\gamma(x, S_x^T(x), d^*(x, S_x^T(x)), u^*(x, S_x^T(x))) \le\ & S_x(x) f(x, l(x), d^*(x, S_x^T(x)) - \tfrac{1}{2}\gamma^2 \|d^*(x, S_x^T(x)\|^2 \\ & + \tfrac{1}{2}\|h(x, l(x))\|^2 \end{aligned}$$

Therefore, by substituting $d = d^*(x, S_x^T(x))$ in (3.23) we obtain

$$K_\gamma(x, S_x^T(x), d^*(x, S_x^T(x)), u^*(x, S_x^T(x))) \le 0.$$

This completes the proof of Theorem 3.2.

[van der Schaft 1996] and [Basar and Bernhard 1991] are two good references of the H_∞ control using the differential game thoery.

3.4 Differential inclusions: The theory of Filippov

Since the differential equations of traffic flow models often carry discontinuous dynamics, we use the theory developed by [Filippov 1964] and [Filippov 1979] to solve the corresponding equations.

Consider a general dynamic control system:

$$\dot{x}(t) = f(t, x, u) \tag{3.24}$$

with function $f : [0, T] \times \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n$ is essentially locally bounded and measurable. The symbol x stands for the state space vector, and u for the input which may be a control or disturbance. The variables u are assumed to be set-valued but bounded. Let us say

$$u \in U(t, x) \tag{3.25}$$

where $U(t, x)$ is a set of functions mapping from $[0, T] \times \mathbf{R}^n$ into some compact subset Ω of $\mathbf{R}^m$ which are measurable in t and continuous in x. Consider

$$F(t, x) = \{f(t, x, u) : u \in U(t, x)\}$$

and then a differential inclusion: $\dot{x} \in F(t, x)$ that reflects the variety of all models (3.24) possible under uncertainty (3.25).

In the classical theory of ordinary differential equations [Coddington and Levinson 1955], solutions exist if f is continuous in x and solutions are unique if f is Lipschitz-continuous in x. In the traffic control applications, however, we shall see that discontinuities in a state-feedback control lead to functions f discontinuous in x, and solutions of (3.24) with a given initial condition x_0 and input u may not exist. To remedy this situation, Filippov introduced a generalized notion of solution as follows:

$$K[f](t,x,u) = \cap_{\delta>0} \cap_{\mu N=0} \overline{\text{co}}\, f(t, B(x,\delta)\backslash N, u), \tag{3.26}$$

where N ranges over all sets of Lebesgue measure (μ) zero, and the notation $\overline{\text{co}}$ denotes closed convex hull.

Definition 3.6 *A vector function $x(\cdot)$ is the solution of (3.24) on the interval $[t_0, t_1]$ in the sense of Filippov if $x(\cdot)$ is absolutely continuous on $[t_0, t_1]$ and for almost all t*

$$\dot{x}(t) \in K[f](t,x,u) \tag{3.27}$$

The main theorem on existence of solutions for a differential inclusion is the idea of Folippov's construction which is to embed $f(t,x,u)$ into a set $K[f](t,x,u)$ as small as possible so that the existence of solutions for the associated differential inclusion is guaranteed. In other words, as described in [Paden and Sastry 1987], *"the content of Filippov's solution is that the tangent vector to a solution where it exists, must lie in a convex closure of the limiting values of the vector field in progressively smaller neighborhoods around the solution point."* We also note that any ambiguities in the definition of f which are confined to a set of measure 0 do not affect $K[f]$.

More recently work about this topic is in [Shevitz and Paden 1994].

3.5 Hybrid system formulation

Systems which have interacting components using more than one modeling paradigm such as difference equations, differential inclusions, boundary dynamics, and finite state machines (FSMs) are called *hybrid systems*. In this thesis we study hybrid systems arising in the control problem of traffic-signal timing for an isolated intersection. These systems are continuous variable, continuous time systems with phased operations. Within each phase the system evolves continuously according to the dynamic law of that phase; when an event (such as queue length changing, special vehicle arrival, accident happening, etc) occurs, the system makes a transition from one phase to another. The signal controller, on the other hand, is acting as a FSM rather than a continuous time system.

The state of the hybrid system is a pair (x, z), where $x \in X$ (the continuous state space) and $z \in Z$ (the discrete state space). In traffic signal control, the discrete state space consists of all types of possible traffic control strategies for the intersection, and the continuous state space is $\mathbf{R^n}$ where n is the total number of approaches or phases for the intersection. That is, we model the controller as an automaton with a finite set of locations and the open loop plant with a continuous state vector $x \in \mathbf{R^n}$, the closed loop system then is a hybrid system.

Chapter 4

An Isolated Two-phase Intersection

In this chapter we consider an isolated two-phase intersection. Isolated implies that we will develop signal control at the intersection without considering adjacent intersections. This is regarded as the starting point of developing a new traffic control strategy. This strategy provides an improved level of control by using better modeling of traffic variables and improved control algorithms.

4.1 A state space model

There are only two traffic flows of vehicles to be served in the intersection shown in Figure 4.1.

State Variables:

x_1: the queue length of the traffic stream in approach A

x_2: the queue length of the traffic stream in approach B

Exogenous inputs:

q_1: the arrival rate of the vehicles at approach A

q_2: the arrival rate of the vehicles at approach B

Parameters:

s_1: the saturation flow rate of approach A

s_2: the saturation flow rate of approach B

L: the total lost time per cycle

Field Control Variables:

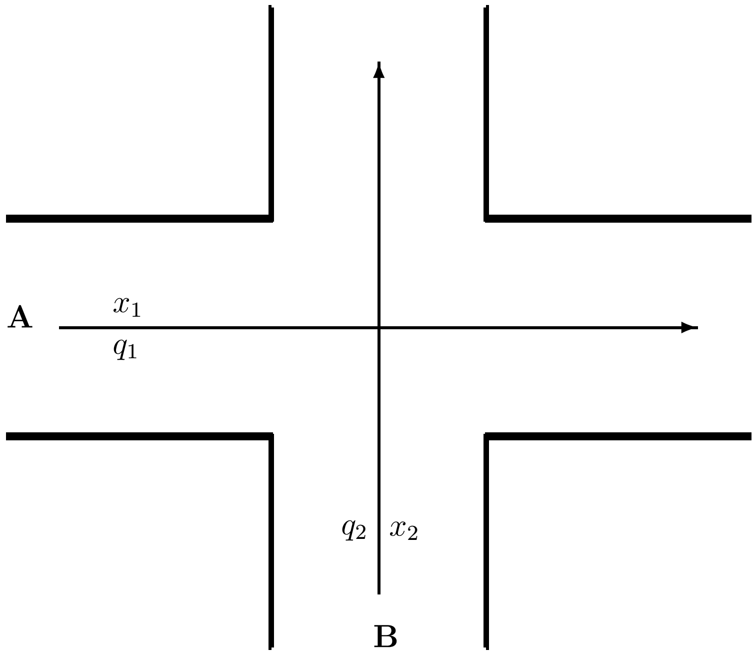

Figure 4.1: A Simple Two-Phase Intersection

g_1: the effective green time on approach A

C: the current average cycle length of the traffic signal.

We assume each approach has a single lane and both s_i to be strictly positive. All other quantities are assumed or constrained to be nonnegative.

The effective green time g_2 on approach B is determined from the effective green time g_1 on approach A and the current cycle time C from the relation

$$g_1 + g_2 = C - L. \tag{4.1}$$

Both effective green times g_1 and g_2 must be nonnegative. Equation (4.1) therefore leads to constraints on the pair of control variables (g_1, C):

$$0 \leq g_1 \leq C - L, \quad C \geq L. \tag{4.2}$$

In other words, the effective green time can be at most equal to the length of the current cycle, and we must impose that the cycle length is at least equal to the lost time L. Conversely, any pair (g_1, C) satisfying these constraints leads to a viable signalization control strategy, although certain boundary points may not fit physically reasonable signalization settings. For example, $g_1 = 0$ corresponds to no green time for direction A, $g_1 = C - L$ corresponds to no effective green time per cycle for direction B, and $C = L$ leads to no effective green time in either direction per cycle.

It will be convenient to introduce a change of control variables:

$$(g_1, C) \to (u_1, u_2) = (\frac{g_1}{C}, 1 - \frac{L}{C} - \frac{g_1}{C}).$$

Instead of g_1 and C, we consider u_1 and u_2 as the independent control variables. In these new coordinates, the constraints (4.2) take the simple triangular form

$$0 \leq u_i, \quad u_1 + u_2 \leq 1. \tag{4.3}$$

This set of control actions (u_1, u_2) is denoted $\mathbf{T}$.

We set up the state dynamic equations based on the following assumption:

Assumption 4.1 *The dynamic rate of the queue length at an approach represents exactly the difference between the arrival density and the flow served at that approach, as long as either the queue length to be served is positive or the arrival density is greater than the flow served. Otherwise, there is no queue dynamics for the approach.*

The flow served at approach i is $s_i \frac{g_i}{C}$ where $i = 1, 2$. Thus, Assumption 4.1 leads to

$$\dot{x}_i = \delta(x_i, q_i - s_i \frac{g_i}{C}), \qquad i = 1, 2 \tag{4.4}$$

where, in general, the function $\delta(\cdot, \cdot)$ is defined by

$$\delta(y, v) = \begin{cases} v & \text{if } y > 0, \text{ or if } y = 0 \text{ and } v > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

In particular, in the situation where both traffic streams are oversaturated (both $x_i > 0$), equation (4.4) assumes the simpler form

$$\dot{x}_i = q_i - s_i \frac{g_i}{C}, \qquad i = 1, 2. \tag{4.6}$$

Division of both sides of (4.1) by C yields

$$\frac{g_1}{C} + \frac{g_2}{C} = 1 - \frac{L}{C}.$$

By the definition of u_2, we have

$$\frac{g_2}{C} = u_2. \tag{4.7}$$

Plugging this equation into equation (4.4), we obtain the following system dynamics:

$$\begin{cases} \dot{x}_1 &= \delta(x_1, q_1 - s_1 u_1) \\ \dot{x}_2 &= \delta(x_2, q_2 - s_2 u_2) \end{cases} \tag{4.8}$$

In the oversaturated case this simplifies to

$$\begin{cases} \dot{x}_1 &= q_1 - s_1 u_1 \\ \dot{x}_2 &= q_2 - s_2 u_2, \end{cases} \tag{4.9}$$

which can be rewritten in matrix format as follows:

$$\dot{x} = q - Gu \tag{4.10}$$

where

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \quad G = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}.$$

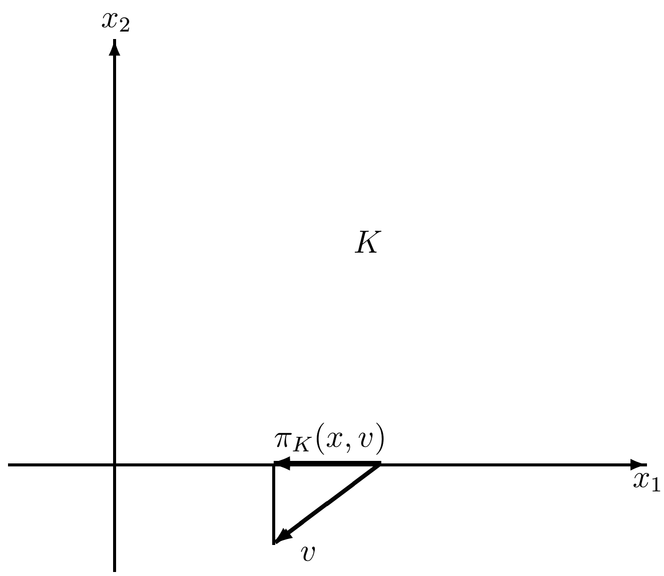

Figure 4.2: Projected Dynamics of the State Space

The system (4.8) can be expressed more compactly if we use the notion of *projected dynamical systems* as in [Dupuis and Nagurney 1993]. Let K be a closed convex subset of $\mathbf{R}^2$. Denote by ∂K and K° the boundary and the interior of K, respectively. Given $x \in \partial K$, we define the inward unit normals to K at x by

$$n(x) = \{\gamma : \|\gamma\| = 1, \langle \gamma, x - y \rangle \leq 0, \forall y \in K\} \tag{4.11}$$

Define the projection map $P_K : \mathbf{R}^2 \to K$ as

$$P_K(x) = \arg\min_{z \in K} \|x - z\|.$$

Given $x \in K$ and $v \in \mathbf{R}^2$, the projection of vector v at x is defined as

$$\pi_K(x, v) = \lim_{\delta \downarrow 0} \frac{P_K(x + \delta v) - x}{\delta}$$

The following properties of π_K were established in [Dupuis 1987]:

Lemma 4.1 (The properties of π_K)

1. *If $x \in K^\circ$, then $\pi_K(x, v) = v$.*
2. *If $x \in \partial K$, then $\pi_K(x, v) = v + \beta(x)n^*(x)$, where*

$$n^*(x) = \arg\max_{n \in n(x)} \langle v, -n \rangle \quad \text{and} \quad \beta(x) = \max\{0, \langle v, -n^*(x) \rangle\}.$$

The projected dynamical system (PDS) proposed in [Dupuis and Nagurney 1993] associated with a (Lipshitz continuous) dynamical system $\dot{x} = F(x)$ and a closed convex set K is

$$\dot{x} = \pi_K(x, F(x)), \quad x(0) = x_0 \in K. \tag{4.12}$$

From Lemma 4.1, we see that the vector field $\pi_K(x, F(x))$ is the same as $F(x)$ when x is in the interior of K, while when x is on the boundary of K we use instead the projection of the vector on K. This ensures that the trajectories of the PDS will always remain in the set K once the initial conditions $x(0) = x_0$ starts in K.

This analysis is easily adapted to the case where the disturbance signal q and control vector u enter into the dynamics. In this case the unprojected system is of the form $\dot{x} = F(x, q, u)$ and the associated PDS (with disturbances and control inputs) with respect to the set K is

$$\dot{x} = \pi_K(x, F(x, q, u)), \quad x(0) = x_0 \in K. \tag{4.13}$$

It is straightforward to check that our traffic control model (4.8) is exactly of this form, with $F(x, q, u) = q - Gu$ and $K = [0, \infty) \times [0, \infty)$, the closed first quadrant. In concise notation, we can write the PDS of (4.8) as

$$\dot{x} = \pi_K(x, q - Gu) \tag{4.14}$$

The existence-uniqueness theory for solutions of a PDS is problematical in the classical ODE theory since the right hand side of the differential equation in (4.13) might not even be continuous in x. This difficulty is addressed by relating (4.12) to a more detailed mechanism for the projected dynamics generally referred to as the Skorokhod problem (see [Dupuis and Nagurney 1993] for details). In this way Dupuis and Nagurney were able to establish that the PDS (4.12) has a unique solution for every $x_0 \in K$, provided K is a polyhedron such as our $K = [0, \infty) \times [0, \infty)$.

4.2 The H_∞ control problem of traffic signalization

We consider the model (4.8) or (4.14) for traffic at an isolated intersection introduced above. The control problem is mathematically much simpler under the assumption that the state vector, which in our case is the set of queue lengths of both traffic streams A and B, is observable at each point in time.

Definition 4.1 **(H_∞ control problem of two-phase intersections)**

Given a tolerance level $\gamma \geq 0$, to design a state-feedback control $x \to u_*(x)$ so that the resulting closed-loop system,

$$\dot{x}(t) = \pi_K(x(t), q(t) - Gu_*(x(t)), \quad x(0) = x_0 \tag{4.15}$$

where q, G, and π_K are defined in section 4.1, satisfies the L_2-gain condition

$$\frac{1}{T}\int_0^T \|x(t)\|^2 \, dt \leq \frac{\gamma^2}{T}\int_0^T \|q(t)\|^2 \, dt + \alpha(x_0)$$

for a nonnegative function $x \to \alpha(x)$ with $\alpha(0) = 0$.

We only consider state vectors x in the first quadrant: both $x_i \geq 0$. The control vector u is limited to the closed and bounded triangle $\mathbf{T}$ given by (4.3). The costate vector $p = (p_1, p_2)^T$ and disturbance vector $q = (q_1, q_2)^T$ may vary throughout $\mathbf{R}^2$.

We are going to apply the approach described in section 3.3 to solve this problem.

First, we compute the desired saddle point (q^*, u^*) of the pre-Hamiltonian function (3.16).

Second, we substitute $q = q^*(x,p)$, $u = u^*(x,p)$ into (3.16) leads to the optimal Hamiltonian

$$H_\gamma^r(x,p) = K_\gamma(x, p, q^*(x,p), u^*(x,p)). \tag{4.16}$$

Next, find a solution $S(x) \geq 0$ of the Hamilton-Jacobi inequality

$$H_\gamma^r(x, S_x^T(x)) \leq 0, \tag{4.17}$$

either a classical solution, or a viscosity sense solution of $H_\gamma^r(x, D^- S^T(x)) \leq 0$. Then $u(x) = u^*(x, S_x^T(x))$ provides a feedback control solving Problem 4.1.

4.3 Solution via H_∞ control theory

There are two sources of discontinuities in H_γ^r to consider in solving (4.17). One is the control constraint $u \in \mathbf{T}$, manifested in (4.27) below. This is fundamental and will be dealt with directly. The other is the discontinuity at $x \in \partial K$ resulting from the projected dynamics. There has been some work in other contexts dealing with this feature directly (see [Dupuis, Ishii and Soner 1990]). We however can use a simpler indirect approach.

We will conduct our analysis using just the Hamiltonian associated with the oversaturated regime, constructing a solution of

$$H_\gamma(x, S_x^T(x)) = 0 \tag{4.18}$$

using the oversaturated Hamiltonian of (4.24). Then we will find that the solution so constructed is in fact a solution of (4.17) for the reflected Hamiltonian as well.

In the oversaturated regime the pre-Hamiltonian may be written out in the more explicit and concise form

$$K_\gamma(x,p,q,u) = p^T(q - Gu) - \frac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2) \tag{4.19}$$

This pre-Hamiltonian is quadratic in q with quadratic term having negative definite coefficient. Hence the maximum in q is easily computed: the critical q is given by $q^*(x,p) = \frac{1}{\gamma^2}p$ and the semi-pre-Hamiltonian $H_{pre,\gamma}(x,p,u) = \max_{q\in\mathbf{R}^2} K_\gamma(x,p,q,u)$ is given by

$$H_{pre,\gamma}(x,p,u) = \frac{1}{2\gamma^2}\|p\|^2 - p^T Gu + \frac{1}{2}\|x\|^2 \tag{4.20}$$

To compute the desired saddle point, it remains to compute arg $\min_{u\in\mathbf{T}} K_\gamma(x,p,q,u)$. Since K_γ is linear in u, this is a simple linear programming problem. In particular, the minimizing $u = u^*(x)$ can always be taken to be one of the vertices of $\mathbf{T}$. The three vertices and their interpretations are:

$u^0 = (0,0)^T$: no green time for either approach, cycle length C equal to the lost time L;

$u^1 = (1,0)^T$: green time for approach A equal to the entire cycle time;

$u^2 = (0,1)^T$: no green time for approach A.

Associated with each vertex is an individual Hamiltonian: $H_{i,\gamma}(x,p) = H_{pre,\gamma}(x,p,u^i)$. Explicitly, we have

$$H_{0,\gamma}(x,p) = \frac{1}{2\gamma^2}\|p\|^2 + \frac{1}{2}\|x\|^2 \tag{4.21}$$

$$H_{1,\gamma}(x,p) = \frac{1}{2\gamma^2}\|p\|^2 - s_1p_1 + \frac{1}{2}\|x\|^2 \tag{4.22}$$

$$H_{2,\gamma}(x,p) = \frac{1}{2\gamma^2}\|p\|^2 - s_2p_2 + \frac{1}{2}\|x\|^2. \tag{4.23}$$

The oversaturated Hamiltonian $H_\gamma(x,p) = \min_{u\in\mathbf{T}} H_{pre,\gamma}(x,p,u)$ is given simply by

$$H_\gamma(x,p) = \min\{H_{0,\gamma}(x,p), H_{1,\gamma}(x,p), H_{2,\gamma}(x,p)\}. \tag{4.24}$$

We should only consider H^r_γ defined for $x \in K$. However we consider H_γ and the $H_{i,\gamma}$ defined for *all* $x \in \mathbf{R}^2$. The Hamiltonians H_γ and H^r_γ agree in K° but not on ∂K. We may complete the squares in $H_{1,\gamma}(x,p)$ and $H_{2,\gamma}(x,p)$ to obtain the alternate expressions

$$H_{1,\gamma}(x,p) = \frac{1}{2\gamma^2}(p_1 - \gamma^2 s_1)^2 + \frac{1}{2\gamma^2}p_2^2 - \frac{\gamma^2}{2}s_1^2 + \frac{1}{2}\|x\|^2, \tag{4.25}$$

$$H_{2,\gamma}(x,p) = \frac{1}{2\gamma^2}p_1^2 + \frac{1}{2\gamma^2}(p_2 - \gamma^2 s_2)^2 - \frac{\gamma^2}{2}s_2^2 + \frac{1}{2}\|x\|^2 \tag{4.26}$$

An immediate observation is that the Hamilton-Jacobi inequality $H_\gamma(x, S_x^T(x)) \leq 0$ can have no global solutions. Indeed observe that $H_{0,\gamma}(x,p) \leq 0$ only for $x = 0, p = 0$. From (4.25) we see that $H_{1,\gamma}(x,p) \leq 0$ has no solution if $\|x\|^2 > \gamma^2 s_1^2$. Similarly, $H_{2,\gamma}(x,p) \leq 0$ has no solution if $\|x\|^2 > \gamma^2 s_2^2$.

We next compute $H_\gamma(x,p)$ explicitly. For this purpose define the following regions in the space of costates:

$\Pi_0 = \{(p_1, p_2)^T : p_1 < 0,\, p_2 < 0\}$,

$\Pi_1 = \{(p_1, p_2)^T : p_1 > 0 \text{ and } s_1 p_1 > s_2 p_2\}$,

$\Pi_2 = \{(p_1, p_2)^T : p_2 > 0 \text{ and } s_1 p_1 < s_2 p_2\}$.

We denote the boundary between Π_1 and Π_2 by

$$\Pi_{12} := \{(p_1, p_2)^T : s_1 p_1 = s_2 p_2,\, p_i > 0\} = \{\rho(s_2, s_1)^T : \rho > 0\}.$$

Then it is easily checked that $u^*(p) = \arg\min_{u \in \mathbf{T}} H_{pre,\gamma}(x, p, u)$ is given by

$$u^*(p) = \begin{cases} u^0 & \text{if } p \in \Pi_0 \\ u^1 & \text{if } p \in \Pi_1 \\ u^2 & \text{if } p \in \Pi_2 \\ \lambda u^1 + (1-\lambda) u^2, \text{ any } 0 \leq \lambda \leq 1 & \text{if } p \in \Pi_{12} \end{cases} \tag{4.27}$$

The Hamiltonian $H_\gamma(x,p)$ is thus given by

$$H_\gamma(x,p) = \begin{cases} H_{1,\gamma}(x,p) & \text{if } p \in \Pi_1 \cup \Pi_{12} \\ H_{2,\gamma}(x,p) & \text{if } p \in \Pi_2 \cup \Pi_{12}. \end{cases} \tag{4.28}$$

The associated Hamilton-Jacobi equation (4.18) can be given piecewise by

$$\begin{aligned} H_{0,\gamma}(x, S_x^T(x)) &= 0 \quad \text{if } S_x^T(x) \in \Pi_0, \\ H_{1,\gamma}(x, S_x^T(x)) &= 0 \quad \text{if } S_x^T(x) \in \Pi_1 \cup \Pi_{12}, \\ H_{2,\gamma}(x, S_x^T(x)) &= 0 \quad \text{if } S_x^T(x) \in \Pi_2 \cup \Pi_{12}. \end{aligned} \tag{4.29}$$

The first step toward constructing a solution $S(x)$ of (4.29) is to postulate that the state-feedback control $u(x) = u^*(S_x^T(x))$ arising from this construction switches between the two values u^1 and u^2 along a half-line Γ:

$$\Gamma = \{(x_1, m x_1)^T : x_1 > 0\}.$$

We shall see later that there is precisely one value of m ($m = s_1/s_2$) for which the construction below succeeds. For now we consider m an unspecified parameter. Denote by X_- the region in the first quadrant strictly below Γ and by X_+ the region in the first quadrant strictly above Γ. Thus we consider the following class of controls:

$$u(x) = \begin{cases} u^1 = (1,0)^T & \text{for } x \in X_-, \\ u^2 = (0,1)^T & \text{for } x \in X_+, \\ (\lambda, 1-\lambda)^T, \text{ some } 0 \leq \lambda \leq 1 & \text{for } x \in \Gamma. \end{cases} \tag{4.30}$$

For $x \in \Gamma$ $u(x)$ is really a convex set of possible controls. The selection of which one to acutally use (i.e. the choice of λ) is related to the Filippov notion of solution of a differential equation with discontinuous right hand side; we will discuss this more in Section 4.5.

The general theory stipulates that the control $u(x)$ be given by that value of u which minimizes $H_{pre,\gamma}(x, p, u)$, i.e.

$$u(x) = u^*(S_x^T(x)), \tag{4.31}$$

with u^* as in (4.27). Under the presumption that rules (4.30) and (4.31) are the same, it follows that along the line Γ, $S_x^T(x)$ should have a value in Π_{12}, the boundary between regions Π_1 and Π_2. For this reason we argue that our solution $S(x)$ of (4.29) should be such that $S_x^T(x_1, mx_1)$ has the form $\rho \cdot (s_2, s_1)^T$ for some $\rho \geq 0$. Hence, when the Hamilton-Jacobi equation (4.29) is restricted to Γ with $S_x^T(x)$ of this special form, we arrive at the equation

$$\begin{aligned} H_{1,\gamma}((x_1, mx_1)^T, \rho(s_2, s_1)^T) &= H_{2,\gamma}((x_1, mx_1)^T, \rho(s_2, s_1)^T) \\ &= \frac{\rho^2}{2\gamma^2}(s_1^2 + s_2^2) - \rho s_1 s_2 + \frac{1}{2}(1+m^2)x_1^2 = 0. \end{aligned} \tag{4.32}$$

Solving this equation for ρ and taking the smaller of the two solutions (since we are looking for the minimal solution of (4.29)) yields

$$\rho = \rho(x_1) := \gamma^2 \cdot \frac{s_1 s_2 - \sqrt{s_1^2 s_2^2 - x_1^2(s_1^2 + s_2^2)(1+m^2)/\gamma^2}}{s_1^2 + s_2^2}. \tag{4.33}$$

The condition

$$|x_1| < x_{1,max} := s_1 s_2 \gamma / \sqrt{(s_1^2 + s_2^2)(1+m^2)}$$

is necessary for a solution to exist. We define Γ_0 to be this segment of Γ:

$$\Gamma_0 = \{(x_1, mx_1)^T : \ 0 \leq x_1 < x_{1,max}\}.$$

We now recover $S(x)$ along Γ_0 by integrating its gradient:

$$\begin{aligned} S(x_1, mx_1) &= \int_0^{x_1} \frac{d}{dy}\{S(y, my)\}\, dy \\ &= \int_0^{x_1} S_x(y, my)(1, m)^T\, dy \\ &= \int_0^{x_1} \rho(y)(s_2 + m s_1)\, dy. \end{aligned} \tag{4.34}$$

Next, since (4.30) postulates that u^1 is the optimal control in X_- we expect $H_{1,\gamma}$ to be the effective form of the Hamiltonian there. We consider the associated Hamiltonian system

$$\begin{array}{rclcl} \dot{x} & = & \frac{\partial^T H_{1,\gamma}}{\partial p}(x,p) & = & \frac{1}{\gamma^2}p - \begin{pmatrix} s_1 \\ 0 \end{pmatrix} \\ \dot{p} & = & -\frac{\partial^T H_{1,\gamma}}{\partial x}(x,p) & = & -x \\ \dot{z} & = & p^T\dot{x} & = & \frac{1}{\gamma^2}p^T p - s_1 p_1 \end{array} \tag{4.35}$$

with initial condition

$$\begin{array}{rcl} x(0) & = & (x_1, m x_1)^T \\ p(0) & = & \rho(x_1)(s_2, s_1)^T, \\ z(0) & = & \int_0^{x_1} \rho(y)(s_2 + m s_1)\, dy \end{array} \tag{4.36}$$

to generate trajectories

$$(x(t,x_1),\ p(t,x_1),\ z(t,x_1)).$$

We hope to cover at least some region $\Omega_- \subseteq X_-$ with such trajectories. This procedure for computing a solution of $H_{1,\gamma}(x, S_x^T(x)) = 0$ in Ω_- amounts to the classical method of characteristics for first order partial differential equations.

Theorem 4.1 *Suppose we can define a smooth function $S(x)$ for $x = (x_1, x_2)^T$ in a neighborhood Ω_- of the line segment $\Gamma_0 = \{(x_1, m x_1)^T : 0 < x_1 < x_{1,max}\}$ implicitly (or in parameterized form) by*

$$S(x) = z(t, x_1) \text{ if } x = x(t, x_1) \tag{4.37}$$

where $(x(t,x_1), p(t,x_1), z(t,x_1))$ are as in (4.35) and (4.36). Then S solves the Hamilton-Jacobi equation $H_{1,\gamma}(x, S_x^T(x)) = 0$ with initial condition $S(x_1, m x_1)$ given by (4.34). Moreover the gradient of S is given implicitly by

$$S_x^T(x) = p(t, x_1) \text{ if } x = x(t, x_1).$$

If in addition the costate vector p stays in region Π_1, Theorem 4.1 gives us a solution S of the Hamilton-Jacobi equation (4.29) in Ω_-. We will see shortly that these hoped for features in the solutions of (4.35) and (4.36) are present when (4.35) is solved backwards in time: $t < 0$. Similarly, starting on Γ_0 and solving the Hamiltonian flow associated with $H_{2,\gamma}$ backwards in time will allow us to construct $S(x)$ for x in a region $\Omega_+ \subseteq X_+$, with the associated gradient $S_x^T(x)$ remaining in region Π_2.

We still need to determine the slope m of the "switching line" Γ. For an initial $x = x_1(1, m)^T \in \Gamma$ and $p = \rho(s_2, s_1)^T \in \Pi_{12}$. The bicharacteristic equations associated with $H_{1,\gamma}$ are

$$\dot{x} = \frac{1}{\gamma^2}p - (s_1, 0)^T, \quad \dot{p} = -x.$$

When integrated in reverse time these should enter the region X_- in state space, and the region Π_1 in costate space. At the initial values this means

$$(-m, 1)\dot{x} \geq 0, \quad (-s_1, s_2)\dot{p} \geq 0.$$

The second condition in particular says

$$-(-s_1, s_2)\, x_1(1, m)^T \geq 0, \quad x_1 \geq 0$$

from which we get that $m \leq s_1/s_2$. Repeating the argument for the bicharacteristics associated with $H_{2,\gamma}$ yields the reverse inequality: $m \geq s_1/s_2$. Thus $m = s_1/s_2$ is necessary for the construction proposed above to work.

Conversely, using $m = s_1/s_2$ the construction described above does succeed in producing a nonnegative C^1 solution of (4.29) in a region $\Omega = \Omega_- \cup \Gamma \cup \Omega_+$ of the first quadrant. We show this theory by experimental verification using sample calculations. Using the software package Mathematica, we can produce graphical plots of various pieces of the solution of the system of ordinary differential equations (4.35) with (4.36). In this way we can view a plot of the graph of S and also check that the costate vector p remains in the appropriate region Π_1 or Π_2.

Figure 4.3: x Plot

We take as an example the case of $s_1 = s_2 = 1$ with $\gamma = 2$. (In this case the slope of the switching line Γ is $m = 1$, as we would expect since the problem is symmetric in x_1 and x_2.) Figure 4.3 shows the state trajectories of (4.35) for $t < 0$, with initial conditions (4.36). We

Figure 4.4: p Plot

see that these trajectories cover a region $\Omega_- \subseteq X_-$, consisting of the triangle with vertices at $(0,0),(1,1),(2,0)$. Figure 4.4 shows that the associated p values do indeed remain in Π_1 as desired. Figure 4.5 plots the values of z (vertically) with respect to (x_1, x_2) (horizontal axes), providing curves on the graph of S. We note in Figure 4.3 that some $x \in \Omega_-$ lie on more than one state trajectory, so that the definition of $S(x)$ is ambiguous. Multivaluedness of S in the context of general Hamilton-Jacobi equations is an interesting phenomenon related to shocks and demanding notions of generalized solutions (such as "viscosity solutions") of the Hamilton-Jacobi equation (see e.g. [James 1993], [Ball and Helton 1996] and [Day 1997]). Such notions are not needed here however. Those x for which $S(x)$ is multivalued involve state-trajectories which have already touched the boundary of the region Ω_-. By truncating these trajectories at the time they contact $\partial\Omega_-$ and only using the truncated trajectories in the construction of Theorem 4.1, S becomes single-valued. This truncation is appropriate for several reasons. Notice in Figure 4.5 that the values of z beyond the truncation point are less than the values for the same x on a trajectory prior to truncation. Only by limiting our construction to truncated trajectories do we obtain a *continuous* function $S(x)$. Secondly, the fact that $S(x)$ is a storage function implies that

$$S(x(0)) - S(x(t)) \leq z(0) - z(t)$$

holds for $t < 0$ along any of the trajectories (4.35), (4.36), even beyond the truncation point, as long as $x(t)$ has not left X_-. Since $z(0) = S(x(0))$ we see that

$$z(t) \leq S(x(t)).$$

This implies that for our construction to be successful we needed to take $S(x)$ to be the largest of the possible values $z(t)$ for $x(t) = x$, which again means discarding the portions of trajectories beyond the truncation point.

This construction produces a solution of $H_\gamma(x, S_x^T(x)) = 0$ in the region Ω_- below Γ in the first quadrant which is covered by the trajectories (4.35), (4.36) associated with $H_{1,\gamma}$ in reverse time. Repeating the construction using $H_{2,\gamma}$ extends the definition to a region

Figure 4.5: Graph of S with two different views

Ω_+ above Γ. Combining the constructions we obtain a solution of (4.29) or (4.18) in $\Omega = \Omega_- \cup \Gamma \cup \Omega_+$. Our solution S is C^1 in Ω because S_x^T is given by the corresponding p values of the two families of trajectories, which agree on Γ. Summing up, we assert the following result.

Theorem 4.2 *The function $S(x)$ constructed as described above, using $m = s_1/s_2$, is a nonnegative C^1 solution of $H_\gamma(x, S_x^T(x)) = 0$, defined for $x \in \Omega = \Omega_- \cup \Gamma \cup \Omega_+$. Along Γ_0 the gradient is given explicitly by*

$$S_x(x_1, mx_1) = \rho(x_1)(s_2, s_1),$$

with $\rho(x_1)$ as in (4.33). More generally the gradient is given implicitly by

$$S_x^T(x) = p(t, x_1) \text{ if } x = x(t, x_1),\ t < 0$$

using the trajectories $x(\cdot, x_1)$ truncated upon first contact with $\partial\Omega$. Moreover, $S(x)$ solves the Hamilton-Jacobi equation associated with the reflected Hamiltonian (4.16):

$$H_\gamma^r(x, S_x^T(x)) = 0, \quad x \in \Omega.$$

We have discussed all the assertions of this theorem except the final one, $H_\gamma(x, S_x^T(x)) = 0$ to $H_\gamma^r(x, S_x^T(x)) = 0$. Theorem 4.2 tells us that $H_\gamma(x, S_x^T(x)) = 0$ in Ω. Since $H_\gamma = H_\gamma^r$ on the interior of Ω, we only need to consider points in $\Omega \cap \partial K$. (We consider the boundary points $x = (x_1, 0)^T$, $x_1 > 0$ to belong to Ω_-, and likewise $x = (0, x_2)^T \in \Omega_+$ for $x_2 > 0$.) For such boundary points we have from Lemma 4.1 that

$$\pi_K(v) = v + \beta n$$

where $\beta \geq 0$ and $n \in n(x)$, the inward unit normals to K at x. It follows then that

$$p^T \pi_K(q - Gu(x)) = p^T(q - Gu(x)) + \beta p^T n$$

for some such β, n. We will see that $S_x(x)n = 0$ on ∂K. From this it will follow that on ∂K

$$S_x \pi_K(q - Gu) = S_x(q - Gu)$$

so that

$$H^r_\gamma(x, S^T_x(x)) = H_\gamma(x, S^T_x(x)) = 0$$

completing our demonstration that $H^r_\gamma(x, S^T_x(x)) = 0$ everywhere in Ω, as claimed. Suppose then that x is in the horizontal portion of ∂K, i.e. $x = (x_1, 0)^T \in \Omega_-$ with $x_1 > 0$. The bicharacteristic equations (4.35) with initial conditions (4.36) used to construct S in Ω_- have a unique solution that reaches $x = x(t)$ at some $t < 0$ without first reaching the diagonal boundary of Ω (at which point we would truncate it), namely the solution with $x(0) = (0, 0)^T$ and $p(0) = \rho(0)(s_2, s_1)^T = (0, 0)^T$, for which we have $x_2(t) = 0$ and $p_2(t) = 0$ for all t. Thus the value of $S^T_x(x)$ will be of the form $p(t) = (p_1(t), 0)^T$ at the time t for which $x(t) = (x_1(t), 0)^T = x$. Since $n = (0, 1)^T$ is the only element of $n(x)$ (unless $x = 0$) we see that indeed $S_x(x)n = 0$. Similar reasoning applies along the vertical side of ∂K. At $x = 0$ we have $S^T_x(0) = 0$ so that$S_x(x)n = 0$ again holds for all $n \in n(0)$.

4.4 The available storage function

The construction of $S(x)$ above is not standard in the current theory of H_∞ problems for smooth systems, as described for instance in the recent monograph [van der Schaft 1996]. The various discontinuities in our hybrid system, due to projected dynamics and our feedback control (4.30), prevent us from invoking the existing theory directly. In spite of this, our $S(x)$ exhibits some important features from the standard theory. Our purpose in this section is to exhibit and demonstrate two of these. In particular we will show that $S(x)$ is the minimal storage function (in Ω) for our PDF (4.14) using control (4.30), and that $(x, S^T_x(x))$, $x \in \Omega$ is an invariant manifold (in a Filippov sense to be explained below) for the Hamiltonian system associated with the oversaturated Hamiltonian (4.24) which the construction was based on. Moreover the trajectories on this manifold all reach the origin in finite time.

Various technical conditions (smoothness of H, hyperbolicity of the Hamiltonian system below, differentiable structure on $\mathcal{X}, \ldots$) allow the construction of a solution to the Hamilton-Jacobi equation $H(x, S^T_x(x)) = 0$ in some domain Ω using the (local) stable invariant manifold (with respect to equilibrium point $(0, 0)$) for the Hamiltonian system associated with

H:

$$\begin{aligned} \dot{x} &= \frac{\partial^T H}{\partial p}(x,p) \\ \dot{p} &= -\frac{\partial^T H}{\partial x}(x,p). \end{aligned}$$

This manifold will (at least in a neighborhood of the origin) be the graph of the gradient of $S(x)$, $x \in \Omega$. That is, the manifold $\{(x, S_x^T(x)) : x \in \Omega\}$ consists of those initial points $(x(0), p(0))$ such that the corresponding solution $(x(t), p(t))$ satisfies $\lim_{t\to\infty}(x(t), p(t)) \to (0,0)$ with $x(t)$ remaining in Ω.

We wish to explore these features for our Hamiltonian H_γ in (4.24) and the input-output system using the control (4.30) generated by our solution. Note that we again deal only with the oversaturated Hamiltonian, rather than the reflected Hamiltonian of (4.16). H_γ^r is discontinuous at $x \in \partial K$, due to the projected dynamics in (4.39). The Hamiltonian system associated with H_γ is

$$\begin{aligned} \dot{x} &= \frac{1}{\gamma^2}p - Gu^i, \quad \text{if } p \in \Pi_i \\ \dot{p} &= -x. \end{aligned} \tag{4.38}$$

We note that the x equation has discontinuities as p moves from one of the regions Π_i to another. Thus we must consider solutions in the sense of Filippov. As we will see, (4.38) has non-unique solutions in this sense, and the solutions associated with the manifold $\mathcal{M} = \{(x, S_x^T(x)) : x \in \Omega\}$ are among them. Thus it is not clear whether we can call $\mathcal{M}$ "*the* stable invariant manifold" for (4.38). We simply make the point that given an initial $x(0) = x \in \Omega$ and $p(0) = S_x^T(x)$ there is a solution of (4.38) with $p(t) = S_x^T(x(t))$ (i.e. remaining on $\mathcal{M}$) and which reaches $x(T) = p(T) = 0$ at a finite $0 < T < \infty$. To this extent at least, $\mathcal{M}$ has features in common with the stable invariant manifold in the smooth theory.

The control associated with our solution S is that of (4.30). Using it in our system (4.14) with the projected dynamics on ∂K, we obtain

$$\dot{x} = \pi_K(x, q - Gu(x)). \tag{4.39}$$

We note that this system is ambiguous for $x \in \Gamma$, since $u(x)$ refers to a convex set of possible values. However when considered in the sense of Filippov, since Γ has Lebesgue measure 0, this ambiguity is irrelevant and we can make sense of (4.39). Our assertion that $S(x)$ is the available storage function in Ω is with reference to this system. Here then is what we will prove.

Theorem 4.3 *Let the function $S(x)$, $x \in \Omega$ be the function constructed as in Theorem 4.1.*

1. *$S(x)$ is the available storage function in Ω for the closed-loop system (4.39). In other words, for any solution of (4.39) with $x(t) \in \Omega$ on $[t_1, t_2]$ we have*

$$S(x(t_2)) - S(x(t_1)) \le \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\,dt,$$

and for any other storage function $\tilde{S}(x)$ we have $S(x) \le \tilde{S}(x)$ for every $x \in \Omega$.

2. *For any initial point $(x, S_x^T(x))$, $x \in \Omega$, (4.38) has a solution, interpreted in the sense of Filippov, defined on an interval $[0, T]$ along with $x(t) \in \Omega$, $p(t) = S_x^T(x(t))$ and with $x(T) = p(T) = 0$.*

Proof We begin with the assertion that S is a storage function in Ω for (4.39). This is an easy consequence of the Hamilton-Jacobi equation $H_\gamma^r(x, S_x^T(x)) = 0$ of Theorem 4.1. Indeed, for $p = S_x^T(x)$ the control (4.31) achieves the infimum over u in (3.17). This and $H_\gamma^r \leq 0$ imply that for any solution of (4.39) with $x(t) \in \Omega$ we have

$$S_x(x(t))x(t) = S_x(x(t))\pi_K(q(t) - Gu(x(t))) \leq \frac{1}{2}(\gamma^2 \|q(t)\|^2 - \|h(x(t))\|^2), \tag{4.40}$$

which of course implies

$$S(x(t_2)) - S(x(t_1)) \leq \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2 \|q(t)\|^2 - \|x(t)\|^2)\, dt.$$

Next we will show that S is minimal among all such storage functions. Our argument is an adaptation of that used for Proposition 7.1.8 in [van der Schaft 1996]. Consider any $x(0) \in \Omega$. We will (below) exhibit a pair $x(t)$, $q(t)$ solving (4.39) on an interval $[0, T]$ with $x(T) = 0$ and for which the storage function inequality becomes an equality:

$$S(x(T)) - S(x(0)) = -S(x(0)) = \int_0^T \frac{1}{2}(\gamma^2 \|q(t)\|^2 - \|x(t)\|^2)\, dt, \tag{4.41}$$

since $S(x(T)) = S(0) = 0$. In light of this, consider any other nonnegative storage function $\tilde{S}$. It follows that

$$-\tilde{S}(x(0)) \leq \tilde{S}(x(T)) - \tilde{S}(x(0)) \leq \int_0^T \frac{1}{2}(\gamma^2 \|q(t)\|^2 - \|x(t)\|^2)\, dt = -S(x(0)).$$

Since $x(0) = x \in \Omega$ was arbitrary, this implies $S \leq \tilde{S}$, establishing our assertion that S is the smallest of all possible storage functions in Ω. The $x(t)$ that we construct for this purpose will be the same as the solution of (4.38) that part 2 of the theorem refers to. We first construct it, then explain the Filippov notion of solution, and then complete the proof of the theorem.

Consider $x \in \Omega_-$ and the solution $x(\cdot), p(\cdot)$ of (4.35) used in our construction of S at x. We can translate the time variable so that $x(0) = x, p(0) = S_x^T(x)$ and $x(t_1) \in \Gamma_0$ at $t_1 > 0$, with $p(t_1) = \rho(x_1(t_1))(s_2, s_1) \in \Pi_{12}$ (the initial condition in (4.36)). Define an associated disturbance by

$$q(t) = \frac{1}{\gamma^2} p(t).$$

Then $x(t), q(t)$ also solve (4.39). (Since $x(t) \in \Omega_-$, the projected dynamics are not involved.) Noting that $q(t)$ achieves the supremum defining $H_{1,\gamma}$ (see the discussion preceeding (4.20)) we find that

$$\begin{aligned} S_x(x(t))x(t) - \tfrac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) &= p(t)(q(t) - Gu(x(t))) - \tfrac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) \\ &= H_{1,\gamma}(x(t), S_x^T(x(t))) \\ &= 0. \end{aligned}$$

In other words, for $x(t), q(t)$ so constructed on $[0, t_1]$, equality is achieved in the dissipation inequality:

$$S(x(t_1)) - S(x(0)) = \int_{t_1}^{0} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\, dt. \tag{4.42}$$

Now we will extend $x(t), p(t)$ to $t > t_1$ using the equations

$$\begin{aligned} \dot{x} &= \tfrac{1}{\gamma^2}p - G\begin{pmatrix} \bar{\lambda} \\ 1 - \bar{\lambda} \end{pmatrix} \\ \dot{p} &= -x \end{aligned} \tag{4.43}$$

with $\bar{\lambda} = \frac{1}{1+m^2}$. This choice of $\bar{\lambda}$ makes the inhomogeneous term $G(\bar{\lambda}, 1-\bar{\lambda})^T = \frac{s_1 s_2}{s_1^2+s_2^2}(s_2, s_1)^T$ parallel to Γ, so that $x(t)$ remains in Γ and $p(t)$ remains in Π_{12} for $t > t_1$. If we write $x(t) = x_1(t)(1, m)^T$ and $p(t) = \rho(t)(s_2, s_1)^T$ in parameterized form, then the Hamiltonian flow (4.43) expressed in terms of $x_1(t)$ and $\rho(t)$ as

$$\begin{aligned} \dot{x}_1(t) &= \tfrac{s_2}{\gamma^2}\rho(t) - \tfrac{s_1}{1+m^2} \\ \dot{\rho}(t) &= -\tfrac{1}{s_2}x_1(t). \end{aligned} \tag{4.44}$$

Note that (4.43) is the Hamiltonian system associated with $H_{pre,\gamma}(x, p, u)$ with $\bar{u} = (\bar{\lambda}, 1 - \bar{\lambda})$ and hence $H_{pre,\gamma}(x, p, \bar{u})$ is constant in t for $t \geq t_1$. Since $p(t) \in \Pi_{12}$, we have that $H_{pre,\gamma}(x(t), p(t), \bar{u}) = H_\gamma(x(t), p(t)) = 0$ and hence $\rho(t)$ is one of the two roots of (4.32). By continuity it follows that $\rho(t) = \rho(x_1(t))$ (where $\rho(x_1)$ is as in (4.33)) and so $p(t) = S_x^T(x(t))$ for $t \geq t_1$.

Note also that the solution $(x(t), p(t))$ reaches $(0, 0)$ at some time T with $t_1 < T < \infty$. To see this, use the fact that $\rho(t) = \rho(x_1(t))$ to write

$$\begin{aligned} \dot{x}_1(t) &= \tfrac{s_2}{\gamma^2}\rho(t) - \tfrac{s_1}{1+m^2} \\ &= -\tfrac{1}{1+m^2}\sqrt{s_1^2 - x_1^2(1+m^2)^2/\gamma^2} \end{aligned}$$

which is strictly negative on any compact subinterval of $[0, x_{max})$. Hence $x(t) = x_1(t)(1, m)^T$ reaches 0 in finite time as asserted. (The choice of $\rho(x_1)$ to be the smaller root of (4.32) is significant here).

Continuing to consider $q(t) = \frac{1}{\gamma^2}p(t)$ as a disturbance on $[t_1, T]$, $x(t)$ continues to be the corresponding solution of (4.39). We are using the value $\bar{\lambda}$ to choose the value of $u(x)$ in (4.30), since $x(t) \in \Gamma$. Again we have

$$S_x(x(t))x(t) - \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) = H_{pre,\gamma}(x(t), p(t)) = H_\gamma(x(t), p(t)) = 0.$$

Thus

$$S(x(T)) - S(x(t_1)) = \int_T^{t_1} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\,dt.$$

Combining this with (4.42) establishes (4.41) used above.

4.5 Discontinuous vector fields and Filippov's solutions

The Hamiltonian system (4.38) and the controlled system (4.39) both have discontinuous dynamics. We interpret these solutions in the sense of Filippov as described in section 3.4.

Consider (4.39). It has discontinuities for $x \in \Gamma$, and

$$K[f](t, x, u) = q(t) - Gu(x)$$

Using what we have already written down in (4.30): for $x \in \Gamma$,

$$K[f] = \{q(t) - (\lambda s_1, (1-\lambda)s_2) : \; 0 \le \lambda \le 1\},$$

the set of convex combinations of $q - Gu^1$ and $q - Gu^2$. If $q(t)$ is parallel to Γ, as it was in our construction above, the opposing directions of the vector fields on opposite sides of Γ imply that the solution of the differential inclusion in this case is unique and remains on Γ. The Filippov solution then corresponds to choosing $u = (\bar{\lambda}, 1-\bar{\lambda})$ with $\bar{\lambda}$ selected so that Gu is parallel to Γ. See Figure 4.6.

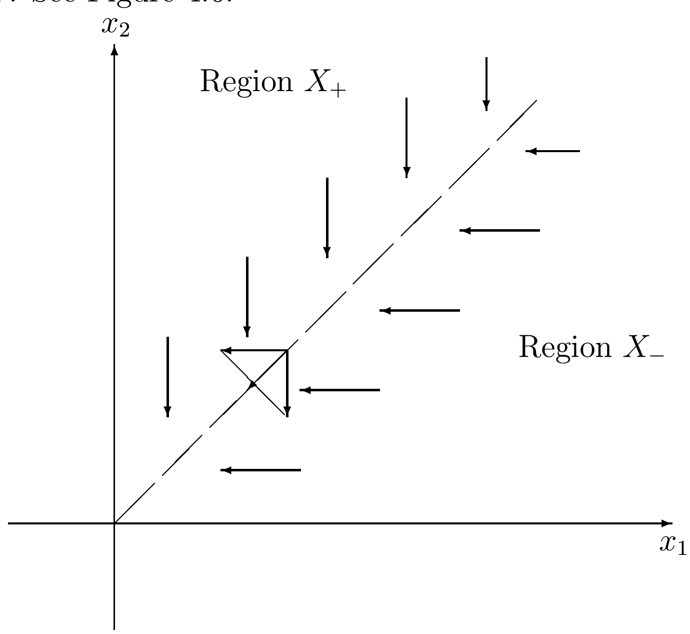

Figure 4.6: Dynamics of the discontinuous flow

This control action can be viewed as an averaging of, or infinitely fast switching between, strategies u^1 and u^2. In actual implementation, this infinite switching frequency cannot be obtained, as there is always a slight delay in implementation of a change of control. Hence, there is some time the trajectory will stay in a region after it crosses Γ, and then again move towards Γ, crossing it once again, and so on. This zig-zag motion is referred to as chattering. This is shown in Figure 4.7. Various modifications of control algorithms have been proposed to mitigate undesirable effects of such chattering.

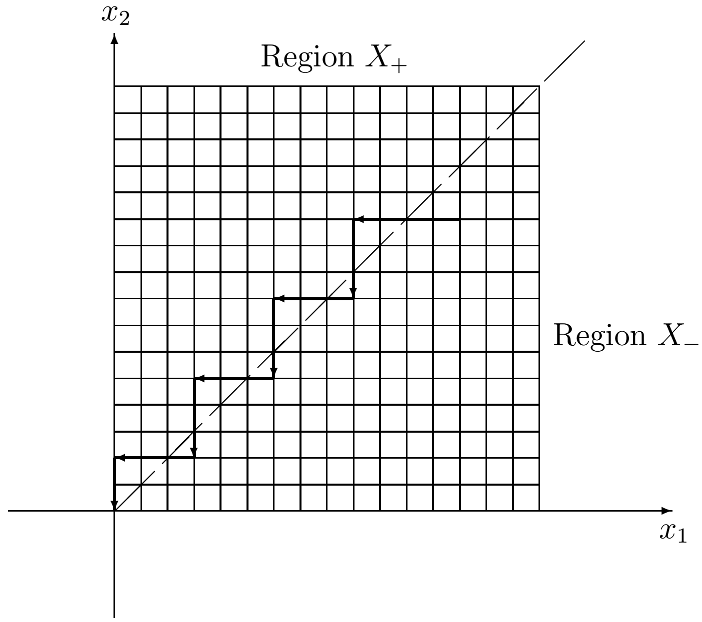

Figure 4.7: Chattering in the Two-phase Intersection Control

The Hamiltonian system (4.38) is rather interesting in the Filippov sense. We have already mentioned that the solutions are not unique, so we consider this system in additional detail. Notice that the discontinuity occurs as a function of the p coordinates, rather the spatial coordinates x as in (4.39).

The behavior of (4.38) can be seen in detail by making an orthonormal change of coordinates in both the x and p planes. Define

$$\mu = (-s_1, s_2) \cdot x/\bar{s} \quad \mu_D = (s_2, s_1) \cdot x/\bar{s}$$
$$\nu = (-s_1, s_2) \cdot p/\bar{s} \quad \nu_D = (s_2, s_1) \cdot p/\bar{s}$$

where $\bar{s} = \sqrt{s_1^2 + s_2^2}$. The coordinates μ_D and ν_D measure the components of x and p in the "diagonal" direction (Γ and Π_{12} respectively), while μ and ν give their components orthogonal to this. In particular $\nu > 0$ indicates $p \in \Pi_2$ so that we should be using $i = 2$ in (4.38). (We are ignoring x in the third quadrant in saying this.) When (4.38) is converted to these cordinates the diagonal components are independent of i:

$$\dot{\mu}_D = \frac{1}{\gamma^2}\nu_D - \frac{s_1 s_2}{\bar{s}}; \quad \dot{\nu}_D = -\mu_D. \tag{4.45}$$

Only the orthogonal components involve the discontinuities:

$$\dot{\mu} = \frac{1}{\gamma^2}\nu + \frac{\pm s_i^2}{\bar{s}}; \quad \dot{\nu} = -\mu. \tag{4.46}$$

Here we use + when $i = 1$ (i.e. for $\nu < 0$) and − when $i = 2$ (i.e. $\nu > 0$). All three of these systems are periodic with period $2\pi\gamma$ having constant angular velocity about their respective centers. The solutions of (4.45) are ellipses centered at $(0, \gamma^2 s_1 s_2/\bar{s})$. The solutions of (4.46) are ellipses centered at $(0, v_i)$ where

$$v_1 = -\gamma^2 s_1^2/\bar{s}, \quad v_2 = +\gamma^2 s_2^2/\bar{s}.$$

The resulting phase portrait for the orthogonal components of (4.38), i.e. (4.46) using $i = 1$ if $\nu < 0$ and $i = 2$ if $\nu > 0$, is illustrated in Figure 4.8. Inspection should convince the reader that solutions are unique, except those passing through the point $\mu = \nu = 0$. We see several solutions through this point: an ellipse in the upper half-plane, an ellipse in the lower half-plane, as well as $\nu(t) \equiv 0, \mu(t) \equiv 0$. This latter is a Filippov solution since the μ component of $K[f]$ at the origin is the interval $[-s_1^2/\bar{s}, s_2^2/\bar{s}]$, which contains 0. In addition $\mu(t), \nu(t)$ can switch from one of these basic soultions to another whenever it passes through the origin.

The solution of (4.43) that we constructed above on $[t_1, T]$ corresponds to $\mu(t) \equiv \nu(t) \equiv 0$ since $x(t) \in \Gamma$ and $p(t) \in \Pi_{12}$. It is then a Filippov solution of (4.38). We now see that there are other solutions of (4.38) starting from the same $x(t_1), p(t_1)$. However, all these alternate solutions will leave the first quadrant (K in x-space) before reaching the origin. To see this, notice that any such solution must have the same diagonal components, $\mu_D(t), \nu_D(t)$ which approach $(0, 0)$ monotonically on $[t_1, T]$. This means that the length of $[t_1, T]$ is at most a quarter period of (4.45). But once a solution of (4.46) leaves $\mu = 0$ it takes exactly a half period until it returns. Thus any solution of (4.38) other than the one we constructed but having the same values $x(t_1), p(t_1)$ will have $\mu(T) \neq 0$ but $\mu_D(T) = 0$. This means the solution has left the first quadrant K. So although (4.38) has many solutions for initial conditions $x(0) \in \Gamma_0$ with $p(0) = S_x^T(x(0))$, there is only one which reaches the origin without first leaving the first quadrant.

We can now complete our proof of part 2 of the theorem. Given $x(0) \in \Omega$ and $p(0) = S_x^T(x(0))$ we constructed $x(t), p(t)$ on $[0, T]$ by solving (4.35) until $x(t_1) \in \Gamma$ and then solving (4.43) until $x(T) = 0$. Since $p(t) \in \Pi_1$ on $[0, t_1]$ and then $p(t) \in \Pi_{12}$ on $[t_1, T]$, we see that we have in fact constructed a solution of (4.38) which maintains $p(t) = S_x^T(x(t))$. (In fact, among the many solutions of (4.38) with these initial values, we have produced the only one which does not leave K before reaching $x(T) = 0$.)

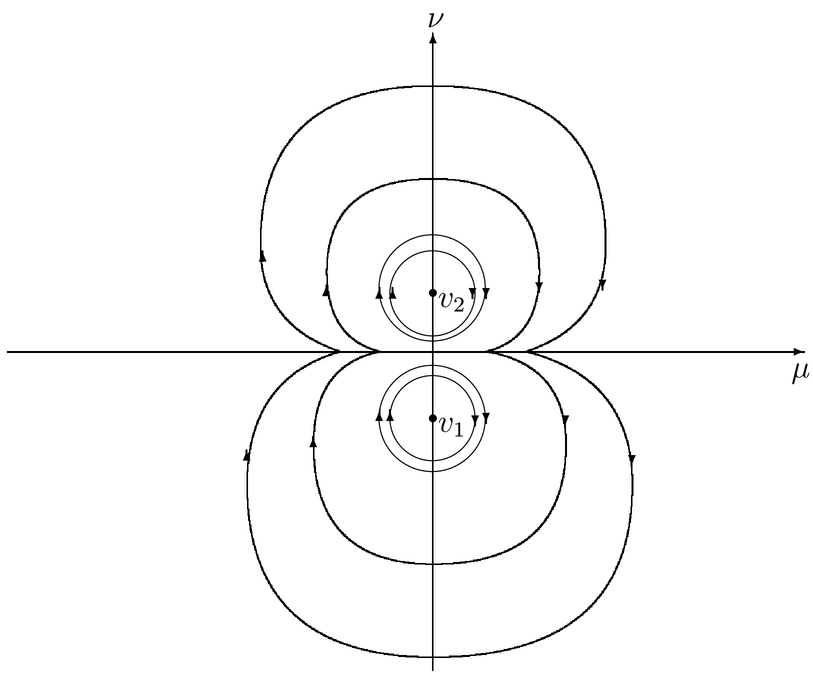

Figure 4.8: Off-diagonal components of Hamiltonian dynamics H_γ

The controlled system (4.39) using (4.30) satisfies the usual criteria of suboptimal H_∞ control: it stabilizes the system with $q(t) \equiv 0$ (Figure 4.6 is the phase portrait) and has the required gain property (this follows from our successful construction of storage function $S(x)$). However the storage function is only defined in a bounded region Ω. Increasing γ will produce a larger Ω, but it will be bounded for each finite γ.

We have worked with the Hamiltonian H_γ of (4.24) in our construction. Our verification of $S(x)$ as the available storage function of (4.39) implicitly involves a second Hamiltonian function, constructed directly from (4.39) and associated with the L_2-gain property of the closed-loop system. Again leaving out the projection dynamics, consider

$$\begin{aligned}\tilde{H}_\gamma(x,p) &= \sup_q \{p^T(q - Gu(x)) - \tfrac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2)\} \\ &= \sup_q K(x,p,q,u(x)),\end{aligned}$$

in terms of the pre-Hamiltonain (4.19). This a distinct Hamiltonian from H_γ. In general it is true that $H_\gamma \leq \tilde{H}_\gamma$, because

$$\begin{aligned} \tilde{H}_\gamma(x,p) &= \sup_q K(x,p,q,u(x)) \\ &\geq \sup_q \inf_{u\in\mathbf{T}} K(x,p,q,u) \ . \\ &= H_\gamma(x,p) \end{aligned}$$

However when $p = S_x^T(x)$ both Hamiltonians agree since the minimizing $u \in \mathbf{T}$ is then given by $u^*(S_x^T(x)) = u(x)$:

$$H_\gamma(x, S_x^T(x)) = \tilde{H}_\gamma(x, S_x^T(x)) = 0.$$

It is actually $\tilde{H}_\gamma(x, S_x^T(x)) \leq 0$ that we used in (4.40) above. One could construct the available storage for (4.39) directly by considering the stable invariant manifold for the Hamiltonian system associated with $\tilde{H}_\gamma$:

$$\begin{aligned} \dot{x} &= \tfrac{1}{\gamma^2}p \ - Gu^i, \quad i = 1(2) \text{ if } x \in X_{-(+)} \\ \dot{p} &= -x. \end{aligned} \tag{4.47}$$

We can again understand the Filippov solutions of this by looking at the orthogonal components μ, ν as before. We now use (4.46) with $i = 1$ for $\mu < 0$ and $i = 2$ for $\mu > 0$ (instead of using the sign of ν as previously). The resulting phase portrait is illustrated in Figure 4.9. Now we see that *all* solutions are unique. Our $(x, S_x^T(x))$, $x \in \Omega$ is again an invariant manifold of solutions, all reaching the origin in finite time. Now however (4.47) has a whole line segment of equilibrium points, connecting v_1 and v_2 along the ν-axis in Figure 4.9. Thus we would not say that the origin is a hyperbolic critical point for (4.47)!

4.6 The implementation of the controller

Switching of green time is based on the criteria of minimizing total system delays by keeping the total average queue lengths of all approaches as small as possible. General implementation of the control law: the green time will always start from g_{min}, after this *initial* period, the control system compares the weighted queue lengths from the two approaches, if the current one is still bigger, then extends the green time by one unit, then compares the queue lengths again, until the other approach has the longer queue or the green time is reaching g_{max}, the system will switch the green into the other approach, and so on. For the optimum throughput, we may set g_{max} relatively big. This will prevent a few stops and save (lost) time for acceleration and clearing. The selection of g_{min} and g_{max} is critical. Too short green phases are impractical or useless, and too long green phases are unacceptable to the stopped drivers of the other approach. Another (not very likely) switching point is that the queue of the other approach is reaching its upper limit, then the green has to be switched over into that approach unless the current approach is also reaching its upper limit. To ensure rapid operation we may set the unit extension (or vehcile interval) as short as practical.

This leads to the following algorithm:

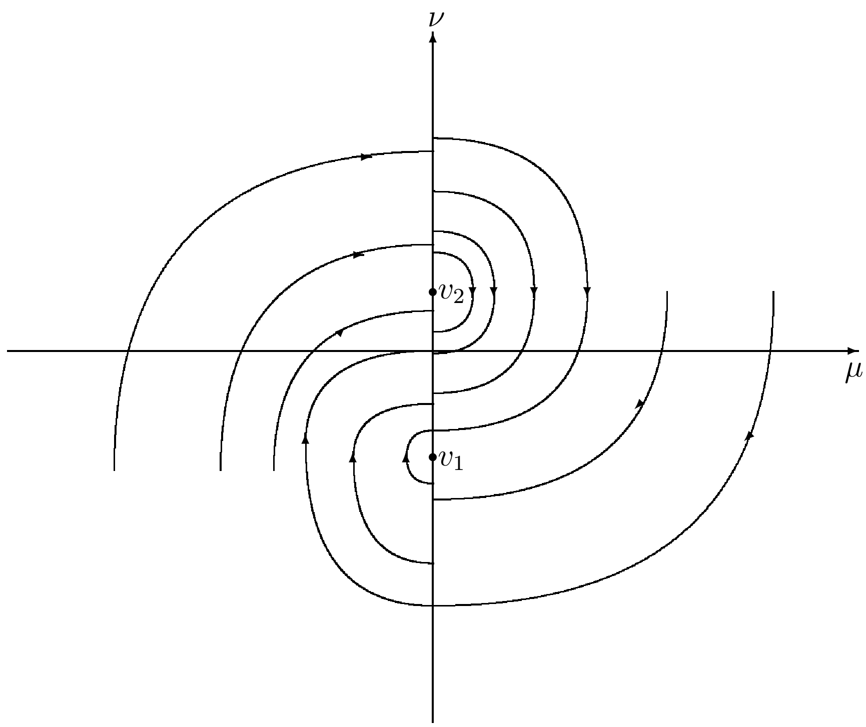

Figure 4.9: Off-diagonal components of alternate Hamiltonian dynamics $\tilde{H}_\gamma$

Algorithm 4.1 *At each instant of time, the controller keeps track of three variables: the two state variables representing queue lengths x_1 and x_2, and the time spent in the current state T_c. Let x_c be the queue length of the approach for which the current signal is green. Let x_o be the queue length of the other approach and x_{max} be the uplimit of x_o. We assume that the minimum green time g_{min}, the unit extension τ, the maximum green time g_{max}, $N = [\frac{g_{max}-g_{min}}{\tau}]$ be the largest integer in the closed interval $[0, \frac{g_{max}-g_{min}}{\tau}]$, the saturation flow rate s_c for the current approach and s_o for the other approach are specified. Then the controller is implemented as follows:*

1. *when $T_c \leq g_{min}$,   stay in the current state.*
2. *when $g_{min} + (k-1)\tau < T_c \leq g_{min} + k\tau < g_{max}$, $(k = 1, 2, \cdots, N)$,   switch state at the end of this interval if $x_o \geq x_{max}$ or $s_c x_c < s_o x_o$ or stay in the current state otherwise.*
3. *when $T_c \geq g_{max}$,   switch state.*

The overall structure of our hybrid system (continuous-time model with FSM controller) is shown in the block diagram of Figure 4.10.

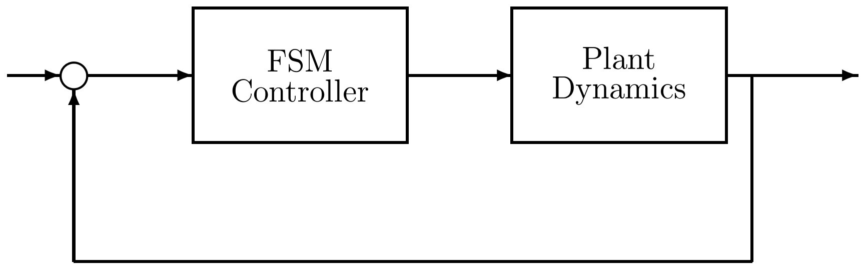

Figure 4.10: Block Diagram for the Hybrid System

One practical necessity is for a *yellow* light as a transition between *green* and *red*. Thus, for approach A, the output of the controller can be green, yellow, or red light, and the same for approach B. Let us define gA, yA and rA, as these output signals for approach A, and gB, yB and rB as those for approach B. There is a relationship between various output signals, such as gA and rB occur together, similarly yA and rB, rA and gB, and rA and yB occur together. That means, there are four states for the control system's FSM and there is a single path of travel between these states. This is shown in Figure 4.11

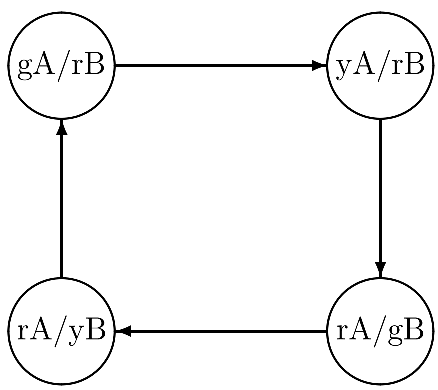

Figure 4.11: FSM for the Simple Two-phase Intersection

The controller's task is to decide when to perform the switch from the current state to the next in the FSM. If we assume a fixed yellow time for both the approaches, then the controller has only two states to worry about, gA/rB and rA/gB. We can consider the fixed yellow times to be approximately modeled as the $g_{min} + \tau$ in the algorithm. Also, fast switching leads to an additional cost in starting-up and clearing time arising from acceleration and deceleration of vehicles. The use of g_{min} is also motivated by the desire to avoid fast switching during service of two long queues of almost equal length.

The other refinement incorporated in the practical algorithm and not immediately arising from consideration of the H_∞ controller is the use of a maximum allowed time g_{max} in either state gA/rB and rA/gB states when both the queues are nonzero. This is included to avoid infinite waiting time for any vehicle, such as the case where there is a saturated flow in one direction at all times, and a relatively short queue of vehicles in the queue in the other direction waiting to be served.

The simple controller we have designed can be easily implemented using a Programmable Array Logic (PAL) sequential device [Shiva 1988] in a mixed signal device as shown in Figure 4.12. The Init signal is for initialization and Clk signal is the clock needed for sequential logic implementation so that the flip flops in the PAL change at the same time after the rising edge of the clock.

Figure 4.12: Hardware for Isolated Signalized Intersection Control

Chapter 5

A Multi-phase Intersection

In general a signalized intersection can have many phases. The controller in such a situation must decide how much time to spend in each phase and also which states to skip, depending on real-time traffic demand. For instance, when there are no vehicles for a left turn lane, the controller may decide to skip the left turn phase. So, at any given time, a signal timing controller has to decide based on the current states whether it needs to change its state, and if the controller needs to change the equations governing its behavior. However, not all state transitions from a given state are permitted. If a transition is to occur, either the next state is completely prescribed, or under certain conditions the controller can skip a state. Each state of the controller FSM generates a distinct closed-loop system represented by a set of ODEs which models the flow of competing traffic approaches at the intersection.

We start this chapter with the same intersection shown in Figure 4.1 from a different aspect of the control variables. To distinguish the model in this chapter from the one in chapter 4, we call the model in chapter 4 as the averaged model.

5.1 A two-phase binary control

Let $\mathcal{U} := \{e_1, e_2\}$ be a standard base of $\mathbf{R}^2$, We introduce a new control variable $u \in \mathcal{U}$. When $u = e_1$, phase ϕ_1 or approach A is green; when $u = e_2$, phase ϕ_2 or approach B is green.

By the same Assumption 4.1 we can obtain the following system dynamics:

$$\begin{cases} \dot{x}_1 &= \delta(x_1, q_1 - \begin{pmatrix} s_1 & 0 \end{pmatrix} u) \\ \dot{x}_2 &= \delta(x_2, q_2 - \begin{pmatrix} 0 & s_2 \end{pmatrix} u) \end{cases} \tag{5.1}$$

where $\delta(\cdot, \cdot)$ is defined in (4.5). In the oversaturated case this simplifies to a matrix form

$$\dot{x} = q - Gu \tag{5.2}$$

where q and G are the same ones as in chapter 4.

As was done for the case of the averaged model, we have the similar PDS, pre-Hamiltonian,semi-pre-Hamiltonian, and Hamiltonian. Finally, we can obtain the similar $u^*(p)$ as for the averaged model (where the minimum of the semi-pre-Hamiltonian was over $u \in \mathbf{T}$ rather than $u \in \mathcal{U}$)

$$u^*(p) = \begin{cases} u_1 & \text{if} \quad p \in \Pi_1 \cup \Pi_{12} \\ u_2 & \text{if} \quad p \in \Pi_2 \cup \Pi_{12} \end{cases} \tag{5.3}$$

where Π_1, Π_2 and Π_{12} are the same ones as in chapter 4.

And we can construct the storage function $S(x)$ in the similar way as in chapter 4, which solve the Hamilton-Jacobi inequality

$$H_\gamma^r(x, S_x^T(x)) \leq 0, \tag{5.4}$$

There also are two sources of discontinuities in H_γ^r in (5.4). The first arises from the control constraint $u \in \mathcal{U}$ that $u(t)$ be in a finite set; thus any nontrivial state-feedback control $x \to u_*(x) = u^*(x, S_x^T(x))$ is necessarily discontinuous as a function of x. The second is the discontinuity along $x_i = 0$ ($i = 1$ or 2) resulting from the form (5.1) of the system equations and the jump discontinuity of the function $(y, v) \to \delta(y, v)$ along $\{(y, v) \colon y = 0, v < 0\}$. This arises from the requirement that the dynamics maintain the physical constraint $x_i \geq 0$ on the components of the state vector x.

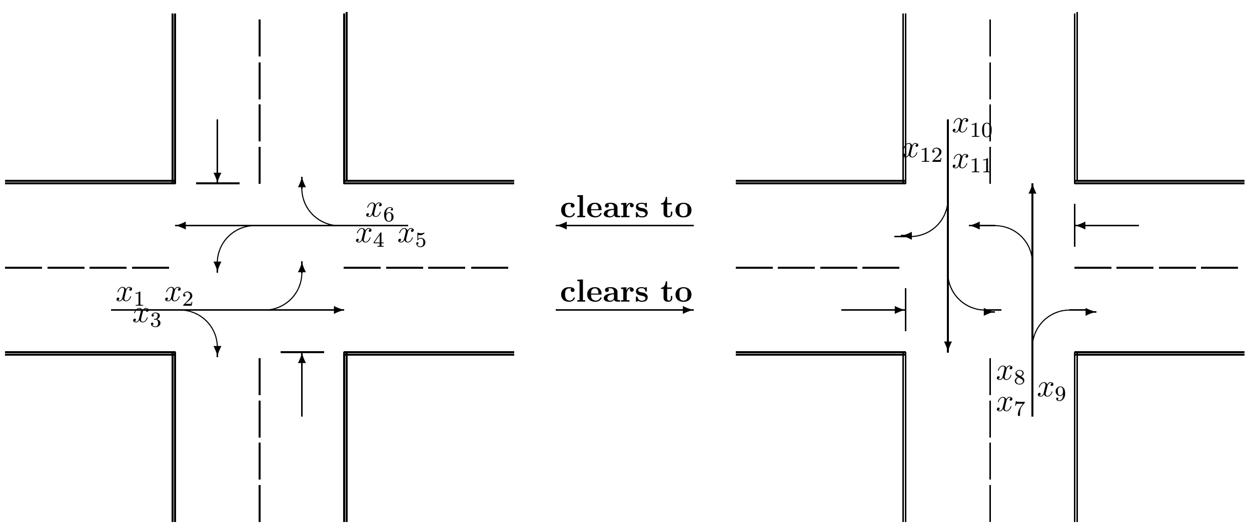

Figure 5.1: Abstract 2-Phase Configuration

Remark: This simple isolated intersection is actually a model for any traffic-signal configuration where the detailed traffic signal control model can be condensed to a 2-phase model.

For example, the configuration in Figure 5.1 leads to the system of differential equations

$$\begin{aligned} \dot{x}_1 &= \delta(x_1, q_1 - (s_1 \;\; 0)u) \\ \dot{x}_2 &= \delta(x_2, q_2 - (s_2 \;\; 0)u) \\ &\vdots \\ \dot{x}_6 &= \delta(x_6, q_6 - (s_6 \;\; 0)u) \\ \dot{x}_7 &= \delta(x_7, q_7 - (0 \;\; s_7)u) \\ &\vdots \\ \dot{x}_{12} &= \delta(x_{12}, q_{12} - (0 \;\; s_{12})u) \end{aligned} \tag{5.5}$$

Introduce aggregate variables $\widetilde{x}_1 = \sum_{i=1}^{6} x_i$, $\widetilde{x}_2 = \sum_{i=7}^{12} x_i$, $\widetilde{q}_1 = \sum_{i=1}^{6} q_i$ and $\widetilde{q}_2 = \sum_{i=7}^{12} q_i$. Set $\widetilde{s}_1 = \sum_{i=1}^{6}$ and $\widetilde{s}_2 = \sum_{i=7}^{12} s_i$. Then system (5.5) can be written in condensed form as

$$\begin{aligned} \dot{\widetilde{x}}_1 &= \delta(\widetilde{x}_1, \widetilde{q}_1 - (\widetilde{s}_1 \;\; 0)u) \\ \dot{\widetilde{x}}_2 &= \delta(\widetilde{x}_2, \widetilde{q}_2 - (0 \;\; \widetilde{s}_2)u). \end{aligned} \tag{5.6}$$

If our control problem is to optimize the total average queue length with respect to the worst case of total average inflow, then we lose nothing by working with the 2-phase system (5.6) rather than the more elaborate but redundant system (5.5). Other more complicated intersections may be reducible to a two state model using what is known as a “workload” formulation in the queueing theory literature.

5.2 An isolated three-phase intersection

A simple three-phase intersection is shown in Figure 5.2. There are three traffic flows of vehicles to be served. Let us denote $I = \{1, 2, 3\}$.

State Variables:

x_i: the queue length of the traffic stream in phase ϕ_i, $\quad i \in I$

Exogenous inputs:

q_i: the arrival rate of the vehicles toward phase ϕ_i, $\quad i \in I$

Parameters:

s_i: the saturation flow rate of phase ϕ_i, $\quad i \in I$

Control Variables:

Let $\{e_1, e_2, e_3\}$ be a standard base of $\mathbf{R}^3$. $u \in \{e_1, e_2, e_3\}$: when $u = e_1$, phase ϕ_1 is green; when $u = e_2$, phase ϕ_2 is green; and when $u = e_3$, phase ϕ_3 is green.

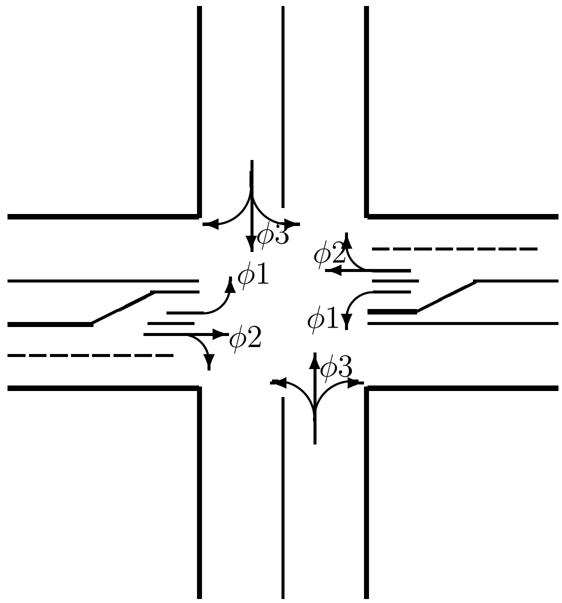

Figure 5.2: Simple Three-phase Signalization

We assume all $s_i\,(i \in I)$ to be strictly positive. All other quantities are assumed or constrained to be nonnegative.

Based on the same Assumption 4.1 we set up the state dynamic equations:

$$\left\{\begin{array}{lcl} \dot{x}_1 & = & \delta(x_1, q_1 - [s_1, 0, 0]u) \\ \dot{x}_2 & = & \delta(x_2, q_2 - [0, s_2, 0]u) \\ \dot{x}_3 & = & \delta(x_3, q_3 - [0, 0, s_3]u) \end{array}\right. \tag{5.7}$$

where $\delta(\cdot, \cdot)$ is defined in (4.5). In the oversaturated case this simplifies to

$$\left\{\begin{array}{lcl} \dot{x}_1 & = & q_1 - [s_1, 0, 0]u \\ \dot{x}_2 & = & q_2 - [0, s_2, 0]u \\ \dot{x}_3 & = & q_3 - [0, 0, s_3]u \end{array}\right. \tag{5.8}$$

Similar to the discussion about projection dynamics in section 4.1, we have the projection dynamics form of model (5.7):

$$\dot{x} = \pi_K(x, q - Gu), \quad x(0) = x_0 \in \Omega. \tag{5.9}$$

where $q = (q_1, q_2, q_3)^T$, $x = (x_1, x_2, x_3)^T$, and Ω is a connected region of the interior of $\mathbf{R}^{3+} = \{x = (x_1, x_2, x_3)^T : \, x_i \geq 0 \, \text{for } i \in I\}$, and,

$$G = \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix}.$$

Following steps similar to those in chapter 4, we form the pre-Hamiltonian

$$K_\gamma(x,p,q,u) = p_1\delta(x_1,q_1-[s_1,0,0]u)+p_2\delta(x_2,q_2-[0,s_2,0]u) \\ +p_3\delta(x_3,q_3-[0,0,s_3]u)-\frac{1}{2}\gamma^2\|q\|^2+\frac{1}{2}\|x\|^2 \quad (5.10)$$

and compute the saddle point $(q^*(x,p),u^*(x,p))$ satisfying

$$K_\gamma(x,p,q,u^*(x,p)) \le K_\gamma(x,p,q^*(x,p),u^*(x,p)) \le K_\gamma(x,p,q^*(x,p),u) \quad (5.11)$$

for all state vectors x, costate vectors p, disturbances q and controls u. We only consider state vectors x in the first quadrant: all $x_i \ge 0$. The PDS dynamics (5.7) maintain this for all solution trajectories. The control vector u is limited to the set $\mathcal{U} := \{e_1,e_2,e_3\}$. The costate vector $p=(p_1,p_2,p_3)^T$ and disturbance vector q may vary throughout $\mathbf{R}^3$.

Substitution of $q=q^*(x,p)$, $u=u^*(x,p)$ into (5.10) leads to the optimal Hamiltonian

$$H_\gamma^r(x,p)=K_\gamma(x,p,q^*(x,p),u^*(x,p)). \quad (5.12)$$

Next, find a solution $S(x)\ge 0$ of the Hamilton-Jacobi inequality

$$H_\gamma^r(x,S_x^T(x)) \le 0, \quad (5.13)$$

Finally we can obtain the desired control

$$u_*(x)=u^*(x,S_x^T(x))$$

which solves the following H_∞ control problem for three-phase intersections.

Definition 5.1 **H_∞ control problem of three-phase intersections** *Given a tolerance level $\gamma \ge 0$, to design a state-feedback control $x \to u_*(x)$ so that the resulting closed-loop system*

$$\dot{x}(t)=\pi_K(x(t),q(t)-Gu_*(x(t)), \quad x(0)=x_0 \quad (5.14)$$

satisfies the L_2-gain condition

$$\frac{1}{T}\int_0^T \|x(t)\|^2\,dt \le \frac{\gamma^2}{T}\int_0^T \|q(t)\|^2\,dt+\alpha(x_0)$$

for a nonnegative function $x\to\alpha(x)$ with $\alpha(0)=0$.

In the oversaturated regime the pre-Hamiltonian may be written out in the more explicit and concise form

$$K_\gamma(x,p,q,u)=p^T(q-Gu)-\frac{1}{2}(\gamma^2\|q\|^2-\|x\|^2) \quad (5.15)$$

This pre-Hamiltonian is quadratic in q with quadratic term having negative definite coefficient. Hence the maximum in q is easily computed: the critical q is given by $q^*(x,p) = \frac{1}{\gamma^2}p$ and the semi-pre-Hamiltonian $H_{pre,\gamma}(x,p,u) = \max_{q\in\mathbf{R}^3} K_\gamma(x,p,q,u)$ is given by

$$H_{pre,\gamma}(x,p,u) = \frac{1}{2\gamma^2}\|p\|^2 - p^T Gu + \frac{1}{2}\|x\|^2 \tag{5.16}$$

To compute the desired saddle point, it remains to compute arg $\min_{u\in\mathcal{U}} K_\gamma(x,p,q,u)$. Let us assign Hamiltonian $H_{1,\gamma}(x,p) = H_{pre,\gamma}(x,p,e_1)$, $H_{2,\gamma}(x,p) = H_{pre,\gamma}(x,p,e_2)$, and $H_{3,\gamma}(x,p) = H_{pre,\gamma}(x,p,e_3)$, that is,

$$\begin{aligned} H_{1,\gamma}(x,p) &= \frac{1}{2\gamma^2}\|p\|^2 - s_1p_1 + \frac{1}{2}\|x\|^2 && (5.17)\\ H_{2,\gamma}(x,p) &= \frac{1}{2\gamma^2}\|p\|^2 - s_2p_2 + \frac{1}{2}\|x\|^2. && (5.18)\\ H_{3,\gamma}(x,p) &= \frac{1}{2\gamma^2}\|p\|^2 - s_3p_3 + \frac{1}{2}\|x\|^2. && (5.19) \end{aligned}$$

The oversaturated Hamiltonian $H_\gamma(x,p) = \min_{u\in\mathcal{U}} H_{pre,\gamma}(x,p,u)$ is given simply by

$$H_\gamma(x,p) = \min\{H_{1,\gamma}(x,p),\ H_{2,\gamma}(x,p),\ H_{3,\gamma}(x,p)\}. \tag{5.20}$$

Now we complete the squares in $H_{1,\gamma}(x,p)$ and $H_{2,\gamma}(x,p)$ to obtain the alternate expressions

$$\begin{aligned} H_{1,\gamma}(x,p) &= \frac{1}{2\gamma^2}(p_1-\gamma^2 s_1)^2 + \frac{1}{2\gamma^2}(p_2^2+p_3^2) - \frac{\gamma^2}{2}s_1^2 + \frac{1}{2}\|x\|^2, && (5.21)\\ H_{2,\gamma}(x,p) &= \frac{1}{2\gamma^2}(p_1^2+p_3^2) + \frac{1}{2\gamma^2}(p_2-\gamma^2 s_2)^2 - \frac{\gamma^2}{2}s_2^2 + \frac{1}{2}\|x\|^2 && (5.22)\\ H_{3,\gamma}(x,p) &= \frac{1}{2\gamma^2}(p_1^2+p_2^2) + \frac{1}{2\gamma^2}(p_3-\gamma^2 s_3)^2 - \frac{\gamma^2}{2}s_3^2 + \frac{1}{2}\|x\|^2 && (5.23) \end{aligned}$$

An immediate observation is that the Hamilton-Jacobi inequality $H_\gamma(x, S_x^T(x)) \le 0$ can have no global solutions. Indeed from (5.21) we see that $H_{1,\gamma}(x,p) \le 0$ has no solution if $\|x\|^2 > \gamma^2 s_1^2$. Similarly, $H_{2,\gamma}(x,p) \le 0$ has no solution if $\|x\|^2 > \gamma^2 s_2^2$, and $H_{3,\gamma}(x,p) \le 0$ has no solution if $\|x\|^2 > \gamma^2 s_3^2$.

We next compute $H_\gamma(x,p)$ explicitly. For this purpose define the following regions in the space of costates:

$$\Pi_1 = \{(p_1,p_2,p_3)^T :\ p_1 > 0 \text{ and } s_1p_1 > \max(s_2p_2, s_3p_3)\},$$

$$\Pi_2 = \{(p_1,p_2,p_3)^T :\ p_2 > 0 \text{ and } s_2p_2 > \max(s_1p_1, s_3p_3)\},$$

$$\Pi_3 = \{(p_1,p_2,p_3)^T :\ p_3 > 0 \text{ and } s_3p_3 > \max(s_1p_1, s_2p_2)\}.$$

We denote the boundary between Π_1, Π_2, and Π_3 by

$$\Pi_{12} := \{(p_1, p_2, p_3)^T : s_1p_1 = s_2p_2 > s_3p_3, \quad p_1, p_2 > 0\},$$

$$\Pi_{13} := \{(p_1, p_2, p_3)^T : s_1p_1 = s_3p_3 > s_2p_2, \quad p_i, p_3 > 0\},$$

$$\Pi_{23} := \{(p_1, p_2, p_3)^T : s_2p_2 = s_3p_3 > s_1p_1, \quad p_2, p_3 > 0\},$$

$$\Pi_{123} := \{(p_1, p_2, p_3)^T : s_1p_1 = s_2p_2 = s_3p_3, \quad p_1, p_2, p_3 > 0\}.$$

Then it is easily checked that $u^*(p) = \arg\min_{u\in\mathcal{U}} H_{pre,\gamma}(x, p, u)$ is given by

$$u^*(p) = \begin{cases} e_1 & \text{if} \quad p \in \Pi_1 \cup \Pi_{12} \cup \Pi_{13} \cup \Pi_{123} \\ e_2 & \text{if} \quad p \in \Pi_2 \cup \Pi_{12} \cup \Pi_{23} \cup \Pi_{123} \\ e_2 & \text{if} \quad p \in \Pi_3 \cup \Pi_{13} \cup \Pi_{23} \cup \Pi_{123} \end{cases} \tag{5.24}$$

The Hamiltonian $H_\gamma(x, p)$ is thus given by

$$H_\gamma(x,p) = \begin{cases} H_{1,\gamma}(x,p) &=& \frac{1}{2}\|x\|^2 + \frac{1}{2\gamma^2}\|p\|^2 - s_1p_1 & \text{if} \quad p \in \Pi_1 \cup \Pi_{12} \cup \Pi_{13} \cup \Pi_{123} \\ H_{2,\gamma}(x,p) &=& \frac{1}{2}\|x\|^2 + \frac{1}{2\gamma^2}\|p\|^2 - s_2p_2 & \text{if} \quad p \in \Pi_2 \cup \Pi_{12} \cup \Pi_{13} \cup \Pi_{123} \\ H_{3,\gamma}(x,p) &=& \frac{1}{2}\|x\|^2 + \frac{1}{2\gamma^2}\|p\|^2 - s_3p_3 & \text{if} \quad p \in \Pi_3 \cup \Pi_{13} \cup \Pi_{23} \cup \Pi_{123} \end{cases} \tag{5.25}$$

The associated Hamilton-Jacobi equation

$$H_\gamma(x, S_x^T(x)) = 0 \tag{5.26}$$

can be given piecewise by

$$\begin{array}{rcll} H_{1,\gamma}(x, S_x^T(x)) &=& 0 & \text{if } S_x^T(x) \in \Pi_1 \cup \Pi_{12} \cup \Pi_{13} \cup \Pi_{123} \\ H_{2,\gamma}(x, S_x^T(x)) &=& 0 & \text{if } S_x^T(x) \in \Pi_2 \cup \Pi_{12} \cup \Pi_{23} \cup \Pi_{123} \\ H_{3,\gamma}(x, S_x^T(x)) &=& 0 & \text{if } S_x^T(x) \in \Pi_3 \cup \Pi_{13} \cup \Pi_{23} \cup \Pi_{123} \end{array}$$

Now we begin to construct a solution $S(x)$ of (5.26). The idea is to decompose Ω into three polyhedrons X_i and the corresponding boundaries so that $S_x^T(x) \in \Pi_i$ whenever $x \in X_i$, where $i \in I$ and construct $S(x)$ by using the method of bicharacteristics of $(x(t), p(t))$ from (x, p) when $t = 0$ to $(0, 0)$ when $t = T$. At the beginning, $x(t) \in X_i, p(t) \in \Pi_i$ and $(x(t), p(t))$ will follow a H_i-bicharacteristic until the moment (t_1) when $x(t)$ reaches an *interface* $\overline{X}_i \cap \overline{X}_j$, $(j \neq i)$ and $p(t)$ also reaches the corresponding interface $\overline{\Pi}_i \cap \overline{\Pi}_j$. After $t > t_1$, $(x(t), p(t))$ will follow a bicharacteristic of the Hamiltonian $H_{ij} = \lambda H_i + (1 - \lambda)H_j$ (which corresponds to a control $u = \lambda e_i + (1 - \lambda)e_j \in \mathcal{U}$) with the parameter λ selected so that $(x(t), p(t))$ will stay in $\overline{(X_i \times \Pi_i)} \cap \overline{(X_j \times \Pi_j)}$ until the moment (t_2) when $x(t)$ reaches the *top interface* $\overline{X}_1 \cap \overline{X}_2 \cap \overline{X}_3$ and $p(t)$ also reaches the corresponding interface $\overline{\Pi}_1 \cap \overline{\Pi}_2 \cap \overline{\Pi}_3$. Finally following a bicharacteristic of the Hamiltonian $H_{123} = \lambda_1 H_1 + \lambda_2 H_2 + (1 - \lambda_1 - \lambda_2)H_3$ at some finite time T, $(x(t), p(t))$ reaches the origin, that is, $x(T) = p(T) = 0$.

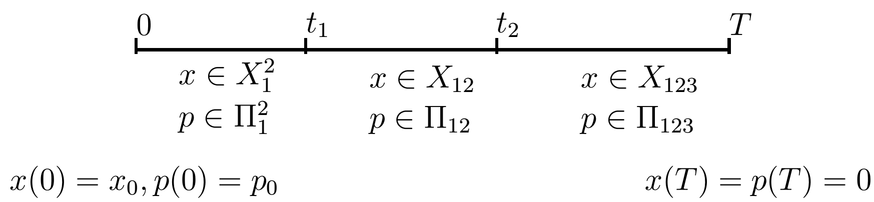

Figure 5.3: The Process of the Bicharacteristics Construction

Let

$$\Pi_i^j = \{(p_1, p_2, p_3)^T : \ s_i p_i > s_j p_j \geq s_k p_k \geq 0\},$$

thus, we have

$$\Pi_i = \Pi_i^j \cup \Pi_i^k$$

where $i \neq j$, $i, j \in I$, and $k \in I \backslash \{i, j\}$.

For example,

$$\Pi_1 = \Pi_1^2 \cup \Pi_1^3$$

where

$$\begin{aligned} \Pi_1^2 &= \{(p_1, p_2, p_3)^T : \ s_1 p_1 > s_2 p_2 \geq s_3 p_3 \geq 0\}, \\ \Pi_1^3 &= \{(p_1, p_2, p_3)^T : \ s_1 p_1 > s_3 p_3 \geq s_2 p_2 \geq 0\}. \end{aligned}$$

In the following construction, we consider $p(t) \in \Pi_1^2$ and the process of construction can be simply shown in Figure 5.3.

Let us denote some constants as follows

$$\begin{aligned} C_{\mu_i} &= s_i, \quad i \in I \\ C_{\nu_{12}} &= \left(\tfrac{1}{s_1^2} + \tfrac{1}{s_2^2}\right)^{-\frac{1}{2}}, \\ C_{\nu_{23}} &= \left(\tfrac{1}{s_2^2} + \tfrac{1}{s_3^2}\right)^{-\frac{1}{2}}, \\ C_{\nu_{13}} &= \left(\tfrac{1}{s_3^2} + \tfrac{1}{s_1^2}\right)^{-\frac{1}{2}}, \quad \text{and} \\ C_{\eta} &= \left(\tfrac{1}{s_1^2} + \tfrac{1}{s_2^2} + \tfrac{1}{s_3^2}\right)^{-\frac{1}{2}} \end{aligned}$$

and some vectors as follows

$$\begin{aligned} \vec{\mu_1} &= C_{\mu_1}(\tfrac{1}{s_1}, 0, 0)^T, \\ \vec{\mu_2} &= C_{\mu_2}(0, \tfrac{1}{s_2}, 0)^T, \\ \vec{\mu_3} &= C_{\mu_3}(0, 0, \tfrac{1}{s_3})^T, \\ \vec{\nu}_{12} &= C_{\nu_{12}}(\tfrac{1}{s_1}, \tfrac{1}{s_2}, 0)^T, \\ \vec{\nu}_{23} &= C_{\nu_{23}}(0, \tfrac{1}{s_2}, \tfrac{1}{s_3})^T, \\ \vec{\nu}_{13} &= C_{\nu_{13}}(\tfrac{1}{s_1}, 0, \tfrac{1}{s_3})^T, \quad \text{and} \\ \vec{\eta} &= C_{\eta}(\tfrac{1}{s_1}, \tfrac{1}{s_2}, \tfrac{1}{s_3})^T. \end{aligned}$$

Then we can choose X_i^j, X_{ij} and X_{123} as follows:

$$\begin{array}{rcl} X_i^j & = & \{a\vec{\eta} + b\vec{\nu}_{ij} + c\vec{\mu}_i : \, a, b \geq 0, \, c > 0\}, \\ & = & \{(x_1, x_2, x_3) : \, s_i x_i > s_j x_j \geq s_k x_k \geq 0\} \\ X_{ij} & = & \{a\vec{\eta} + b\vec{\nu}_{ij} : \, a, b \geq 0\}, \quad \text{and} \\ X_{123} & = & \{a\vec{\eta} : \, a \geq 0\} \end{array}$$

where $i \neq j$, $i, j \in I$, and $k \in I \backslash \{i, j\}$.

Let $\sum_{i=1}^{3} \lambda_i = 1, \quad \lambda_i \geq 0, i \in I$, and

$$\begin{array}{rcl} H_i & = & \frac{1}{2\gamma^2}\|p\|^2 - p^T G e_i + \frac{1}{2}\|x\|^2, \quad i \in I \\ H_{123} & = & \sum_{i=1}^{3} \lambda_i H_i \\ & = & \frac{1}{2\gamma^2}\|p\|^2 - p^T G u_{123}^* + \frac{1}{2}\|x\|^2 \end{array}$$

where $u_{123}^* = (\lambda_1, \lambda_2, \lambda_3)^T$ is chosen so that $X_{123} \times \Pi_{123}$ is invariant for the Hamiltonian system:

$$\begin{array}{rclcl} \dot{x} & = & \frac{\partial^T H_{123}}{\partial p}(x, p) & = & \dfrac{1}{\gamma^2} p - G u_{123}^* \\ \dot{p} & = & -\frac{\partial^T H_{123}}{\partial x}(x, p) & = & -x \end{array} \tag{5.27}$$

In other words, we want

$$G u_{123}^* = C\vec{\eta}$$

where C is a constant. That is,

$$(s_1\lambda_1, s_2\lambda_2, s_3\lambda_3)^T = C C_\eta (\frac{1}{s_1}, \frac{1}{s_2}, \frac{1}{s_3})^T$$

$$\lambda_i = C C_\eta \frac{1}{s_i^2}$$

By $\sum_{i=1}^{3} \lambda_i = 1$,

$$1 = \sum_{i=1}^{3} C C_\eta \frac{1}{s_i^2} = C C_\eta \sum_{i=1}^{3} \frac{1}{s_i^2}$$

By the definition of C_η, we obtain

$$C C_\eta \frac{1}{C_\eta^2} = 1$$

So

$$C = C_\eta$$

$$\lambda_i = \frac{C_\eta^2}{s_i^2}.$$

Therefore, we solve

$$H_{123} = \frac{1}{2\gamma^2}\|p\|^2 - C_\eta p^T \vec{\eta} + \frac{1}{2}\|x\|^2, \tag{5.28}$$

and the associated system:

$$\begin{aligned} \dot{x} &= \frac{1}{\gamma^2}p - C_\eta \vec{\eta} \\ \dot{p} &= -x \end{aligned} \tag{5.29}$$

In $X_{123} \times \Pi_{123}$ we may write

$$\begin{cases} x(t) &= a(t)\vec{\eta} \\ p(t) &= \gamma\alpha(t)\vec{\eta} \\ x(T) &= p(T) = 0 \end{cases} \tag{5.30}$$

where $a(t)$ and $\alpha(t)$ satisfy the following equations:

$$\begin{cases} \dot{a}(t) &= \frac{1}{\gamma}\alpha(t) - C_\eta \\ \dot{\alpha}(t) &= -\frac{1}{\gamma}a(t) \\ a(T) &= \alpha(T) = 0 \end{cases} \tag{5.31}$$

Clearly, the above equation has the equilibrium point: $\alpha = \gamma C_\eta, a = 0$. Since the matrix $\begin{bmatrix} 0 & \frac{1}{\gamma} \\ -\frac{1}{\gamma} & 0 \end{bmatrix}$ has the eigenvalues $i\frac{1}{\gamma}$ and $-i\frac{1}{\gamma}$, we have the solution of the equation (5.31):

$$\begin{cases} a(t) &= -\gamma C_\eta sin(\frac{t-T}{\gamma}) \\ \alpha(t) &= \gamma C_\eta(1 - cos(\frac{t-T}{\gamma})) \end{cases} \tag{5.32}$$

That is, the solution $(a(t), \alpha(t))$ rotates clockwisely around $(0, \gamma C_\eta)$ in the (a, α)-plane with the period $2\gamma\pi$.

Consider the equation

$$H_{123}(\xi\vec{\eta}, \rho\vec{\eta}) = 0. \tag{5.33}$$

By (5.28), we have

$$\frac{1}{2\gamma^2}\rho^2 - \rho C_\eta + \frac{1}{2}\xi^2 = 0$$

We can solve this equation for $\rho(\xi)$ as

$$\rho(\xi) = \gamma^2(C_\eta \pm \sqrt{C_\eta^2 - \frac{\xi^2}{\gamma^2}})$$

where $\xi < C_\eta \gamma \equiv \xi_{\max}$. Let us take the smaller one $\rho(\xi) = \gamma^2(C_\eta - \sqrt{C_\eta^2 - \frac{\xi^2}{\gamma^2}})$, thus we can obtain

$$\begin{aligned} S(l\vec{\eta}) &= \int_0^l \frac{d}{d\xi}\{S(\xi\vec{\eta})\}d\xi \\ &= \int_0^l S_x(\xi\vec{\eta})\vec{\eta}d\xi \\ &= \int_0^l \rho(\xi)\vec{\eta}^T\vec{\eta}d\xi \\ &= \int_0^l \rho(\xi)d\xi \end{aligned} \tag{5.34}$$

Next let $H_{12} = \lambda_1 H_1 + \lambda_2 H_2$ where $\lambda_i \geq 0$, $\lambda_1 + \lambda_2 = 1$. That is,

$$H_{12} = \frac{1}{2\gamma^2}\|p\|^2 - p^T G u_{12}^* + \frac{1}{2}\|x\|^2$$

where $u_{12}^* = (\lambda_1, \lambda_2, 0)^T$ is chosen so that $X_{12} \times \Pi_{12}$ is invariant for the Hamiltonian system:

$$\begin{aligned} \dot{x} &= \tfrac{\partial^T H_{12}}{\partial p}(x,p) &&= \frac{1}{\gamma^2}p - G u_{12}^* \\ \dot{p} &= -\tfrac{\partial^T H_{12}}{\partial x}(x,p) &&= -x \end{aligned} \tag{5.35}$$

Thus, we want $Gu_{12}^* = (\lambda_1 s_1, \lambda_2 s_2, 0)^T$ and must be a nonnegative combination of $\vec{\eta}$ and $\vec{\nu}_{12}$. We notice that the third coordinate is 0, thus we may set

$$Gu_{12}^* = CC_{\nu_{12}}(\frac{1}{s_1}, \frac{1}{s_2}, 0)$$

$$\lambda_i = C\frac{C_{\nu_{12}}}{s_i^2}, \quad i = 1, 2$$

By $1 = \lambda_1 + \lambda_2$ and the definition of $C_{\nu_{12}}$, we have

$$C\frac{C_{\nu_{12}}}{C_{\nu_{12}}^2} = 1$$

Thus,

$$C = C_{\nu_{12}} \quad \text{and} \quad Gu_{12}^* = C_{\nu_{12}}\vec{\nu}_{12}$$

Therefore, we solve

$$H_{12} = \frac{1}{2\gamma^2}\|p\|^2 - C_{\nu_{12}} p^T \vec{\nu}_{12} + \frac{1}{2}\|x\|^2,$$

and the associated Hamiltonian system is

$$\begin{aligned} \dot{x} &= \frac{1}{\gamma^2}p - C_{\nu_{12}}\vec{\nu}_{12} \\ \dot{p} &= -x \end{aligned} \tag{5.36}$$

We seek a solution in $[t_1, t_2]$ in $X_{12} \times \Pi_{12}$ which reaches $X_{123} \times \Pi_{123}$ at time t_2. We may write

$$\left\{ \begin{aligned} x(t) &= a(t)\vec{\eta} + b(t)\vec{\nu}_{12} \\ p(t) &= \gamma(\alpha(t)\vec{\eta} + \beta(t)\vec{\nu}_{12}) \\ b(t_2) &= \beta(t_2) = 0 \end{aligned} \right. \tag{5.37}$$

Then the Hamiltonian system is

$$\begin{aligned} \dot{a}(t)\vec{\eta} + \dot{b}(t)\vec{\nu}_{12} &= \frac{1}{\gamma}\alpha(t)\vec{\eta} + (\frac{1}{\gamma}\beta(t) - C_{\nu_{12}})\vec{\nu}_{12} \\ \gamma(\dot{\alpha}(t)\vec{\eta} + \dot{\beta}(t)\vec{\nu}_{12}) &= -a(t)\vec{\eta} - b(t)\vec{\nu}_{12} \end{aligned} \tag{5.38}$$

Thus, we have the following equations for $a(t), \alpha(t), b(t)$ and $\beta(t)$:

$$\begin{cases} \dot{a}(t) &= \frac{1}{\gamma}\alpha(t) \\ \dot{\alpha}(t) &= -\frac{1}{\gamma}a(t) \\ a(t_2) &= -\gamma C_\eta \sin\frac{t_2-T}{\gamma} \\ \alpha(t_2) &= \gamma C_\eta (1-\cos\frac{t_2-T}{\gamma}) \end{cases} \tag{5.39}$$

where we use (5.32) for the initial values $a(t_2), \alpha(t_2)$. And

$$\begin{cases} \dot{b}(t) &= \frac{1}{\gamma}\beta(t) - C_{\nu_{12}} \\ \dot{\beta}(t) &= -\frac{1}{\gamma}b(t) \\ b(t_2) &= \beta(t_2) = 0 \end{cases} \tag{5.40}$$

Solving (5.39) we get

$$\begin{cases} a(t) &= a(t_2)\cos(\frac{t-t_2}{\gamma}) + \alpha(t_2)\sin(\frac{t-t_2}{\gamma}) \\ \alpha(t) &= -a(t_2)\sin(\frac{t-t_2}{\gamma}) + \alpha(t_2)\cos(\frac{t-t_2}{\gamma}) \end{cases} \tag{5.41}$$

It can be simplified as

$$\begin{cases} a(t) &= \gamma C_\eta (\sin(\frac{t-t_2}{\gamma}) - \sin(\frac{t-T}{\gamma})) \\ \alpha(t) &= \gamma C_\eta (\cos(\frac{t-t_2}{\gamma}) - \cos(\frac{t-T}{\gamma})) \end{cases} \tag{5.42}$$

Solving (5.40) we get

$$\begin{cases} b(t) &= -\gamma C_{\nu_{12}} sin(\frac{t-t_2}{\gamma}) \\ \beta(t) &= \gamma C_{\nu_{12}} (1 - cos(\frac{t-t_2}{\gamma})) \end{cases} \tag{5.43}$$

These solutions can be described in Figure 5.4, where $(a(t), \alpha(t))$ rotates clockwisely around the origin with the period $2\gamma\pi$; and $(b(t), \beta(t))$ rotates clockwisely around $(0, \gamma C_{\nu_{12}})$ with the same period.

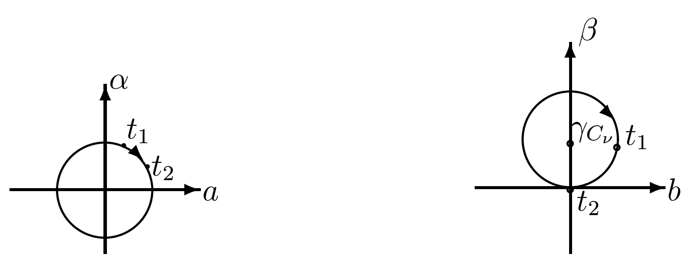

Figure 5.4: A Decoupled Clockwise Rotations

Finally we want to solve the H_1 system so that $X_1 \times \Pi_1$ is invariant under u_1^*. Note that $Gu_1^* = (s_1, 0, 0)^T = C_{\mu_1}\vec{\mu_1}$, so we solve

$$H_1 = \frac{1}{2\gamma^2}\|p\|^2 - C_{\mu_1} p^T \vec{\mu_1} + \frac{1}{2}\|x\|^2$$

and the associated Hamiltonian system is

$$\begin{aligned} \dot{x} &= \frac{1}{\gamma^2}p - C_{\mu_1}\vec{\mu_1} \qquad x(0) = x_0 \\ \dot{p} &= -x \end{aligned} \tag{5.44}$$

where $x(t_1) \in X_{12}$ and $p(t_1) \in \Pi_{12}$. Similarly to the above procedure, we write that

$$\begin{cases} x(t) &= a(t)\vec{\eta} + b(t)\vec{\nu}_{12} + c(t)\vec{\mu_1} \\ p(t) &= \gamma(\alpha(t)\vec{\eta} + \beta(t)\vec{\nu}_{12} + \theta(t)\vec{\mu_1}) \\ c(t_1) &= \theta(t_1) = 0 \end{cases} \tag{5.45}$$

By (5.44),

$$\begin{aligned} \dot{a}(t)\vec{\eta} + \dot{b}(t)\vec{\nu}_{12} + \dot{c}(t)\vec{\mu_1} &= \frac{1}{\gamma}\alpha(t)\vec{\eta} + \frac{1}{\gamma}\beta(t)\vec{\nu}_{12} + (\frac{1}{\gamma}\theta(t) - C_{\mu_1})\vec{\mu_1} \\ \gamma(\dot{\alpha}(t)\vec{\eta} + \dot{\beta}(t)\vec{\nu}_{12} + \dot{\theta}(t)\vec{\mu_1}) &= -a(t)\vec{\eta} - b(t)\vec{\nu}_{12} - c(t)\vec{\mu_1} \end{aligned} \tag{5.46}$$

Thus, we have the following equations for $a(t), \alpha(t), b(t), \beta(t), c(t)$ and $\theta(t)$:

$$\begin{cases} \dot{a}(t) &= \frac{1}{\gamma}\alpha(t) \\ \dot{\alpha}(t) &= -\frac{1}{\gamma}a(t) \\ a(t_1) &= \gamma C_\eta(\sin\frac{t_1-t_2}{\gamma} - \sin\frac{t_1-T}{\gamma}) \\ \alpha(t_1) &= \gamma C_\eta(\cos\frac{t_1-t_2}{\gamma} - \cos\frac{t_1-T}{\gamma}) \end{cases} \tag{5.47}$$

$$\begin{cases} \dot{b}(t) &= \frac{1}{\gamma}\beta(t) \\ \dot{\beta}(t) &= -\frac{1}{\gamma}b(t) \\ b(t_1) &= -\gamma C_{\nu_{12}}\sin\frac{t_1-t_2}{\gamma} \\ \beta(t_1) &= \gamma C_{\nu_{12}}(1 - \cos\frac{t_1-t_2}{\gamma}) \end{cases} \tag{5.48}$$

and

$$\begin{cases} \dot{c}(t) &= \frac{1}{\gamma}\theta(t) - C_{\mu_1} \\ \dot{\theta}(t) &= -\frac{1}{\gamma}c(t) \\ c(t_1) &= \theta(t_1) = 0 \end{cases} \tag{5.49}$$

Solving (5.47) we get

$$\begin{cases} a(t) &= a(t_1)\cos(\frac{t-t_1}{\gamma}) + \alpha(t_1)\sin(\frac{t-t_1}{\gamma}) \\ \alpha(t) &= -a(t_1)\sin(\frac{t-t_1}{\gamma}) + \alpha(t_1)\cos(\frac{t-t_1}{\gamma}) \end{cases} \tag{5.50}$$

It can be simplified as

$$\begin{cases} a(t) &= \gamma C_\eta(\sin(\frac{t-t_2}{\gamma}) - \sin(\frac{t-T}{\gamma})) \\ \alpha(t) &= \gamma C_\eta(\cos(\frac{t-t_2}{\gamma}) - \cos(\frac{t-T}{\gamma})) \end{cases} \tag{5.51}$$

Solving (5.48) we get

$$\begin{cases} b(t) &= \gamma C_{\nu_{12}}(\sin(\frac{t-t_1}{\gamma}) - \sin(\frac{t-t_2}{\gamma})) \\ \beta(t) &= \gamma C_{\nu_{12}}(\cos(\frac{t-t_1}{\gamma}) - \cos(\frac{t-t_2}{\gamma})) \end{cases} \tag{5.52}$$

Solving (5.49) we get

$$\begin{cases} c(t) &= -\gamma C_{\mu_1} sin(\frac{t-t_1}{\gamma}) \\ \theta(t) &= \gamma C_{\mu_1}(1 - cos(\frac{t-t_1}{\gamma})) \end{cases} \tag{5.53}$$

Again the solutions $(a(t), \alpha(t))$ and $(b(t), \beta(t))$ rotate around the origin with the period $2\gamma\pi$ and $(c(t), \theta(t))$ rotates around $(0, \gamma C_{\mu_1})$ with the same period.

Let $\Omega(\gamma) \subset \Omega$ be the set of $x(0)$'s which can be achieved from the origin by this construction.

Given $x(0) = x_0 \in \Omega(\gamma) \cap X_1^2$ we have prescribed values $a(0) = a_0$, $b(0) = b_0$, and $c(0) = c_0$. and then we can solve T, t_2, t_1 and $\alpha(t), \beta(t), \theta(t)$ as follows.

By (5.53), we have

$$t_1 = \gamma \sin^{-1}(\frac{c_0}{\gamma C_{\mu_1}}), \quad \text{and}$$

$$\theta(t) = \gamma C_{\mu_1}(1 - cos(\frac{1}{\gamma}t - \sin^{-1}(\frac{c_0}{\gamma C_{\mu_1}}))).$$

By (5.52), we have

$$t_2 = \gamma \sin^{-1}(\frac{b_0}{\gamma C_{\nu_{12}}} + \frac{c_0}{\gamma C_{\mu_1}}), \quad \text{and}$$

$$\beta(t) = \gamma C_{\nu_{12}}(cos(\frac{1}{\gamma}t - \sin^{-1}(\frac{c_0}{\gamma C_{\mu_1}})) - cos(\frac{1}{\gamma}t - \sin^{-1}(\frac{b_0}{\gamma C_{\nu_{12}}} + \frac{c_0}{\gamma C_{\mu_1}}))).$$

By (5.51), we have

$$T = \gamma \sin^{-1}(\frac{a_0}{\gamma C_\eta} + \frac{b_0}{\gamma C_{\nu_{12}}} + \frac{c_0}{\gamma C_{\mu_1}}), \quad \text{and}$$

$$\alpha(t) = \gamma C_\eta(cos(\frac{1}{\gamma}t - \sin^{-1}(\frac{b_0}{\gamma C_{\nu_{12}}} + \frac{c_0}{\gamma C_{\mu_1}})) - cos(\frac{1}{\gamma}t - \sin^{-1}(\frac{a_0}{\gamma C_\eta} + \frac{b_0}{\gamma C_{\nu_{12}}} + \frac{c_0}{\gamma C_{\mu_1}}))).$$

All above formulas make sense if and only if

$$\frac{a_0}{\gamma C_\eta} + \frac{b_0}{\gamma C_{\nu_{12}}} + \frac{c_0}{\gamma C_{\mu_1}} \leq 1$$

that is,

$$\frac{a_0}{C_\eta} + \frac{b_0}{C_{\nu_{12}}} + \frac{c_0}{C_{\mu_1}} \leq \gamma$$

Thus,

$$\Omega(\gamma) \cap X_1^2 = \{a_0\vec{\eta} + b_0\vec{\nu}_{12} + c_0\vec{\mu_1} \in \Omega : \frac{a_0}{C_\eta} + \frac{b_0}{C_{\nu_{12}}} + \frac{c_0}{C_{\mu_1}} \leq \gamma\}.$$

Similar analysis can carry over to other regions, we also have

$$\Omega(\gamma) \cap X_1^3 = \{a_0\vec{\eta} + b_0\vec{\nu}_{13} + c_0\vec{\mu_1} \in \Omega : \frac{a_0}{C_\eta} + \frac{b_0}{C_{\nu_{13}}} + \frac{c_0}{C_{\mu_1}} \leq \gamma\},$$

$$\Omega(\gamma) \cap X_2^3 = \{a_0\vec{\eta} + b_0\vec{\nu}_{23} + c_0\vec{\mu_2} \in \Omega : \frac{a_0}{C_\eta} + \frac{b_0}{C_{\nu_{23}}} + \frac{c_0}{C_{\mu_2}} \leq \gamma\},$$

etc.

In general, we have the explicit form of $\Omega(\gamma)$:

$$\Omega(\gamma) = \cup_{1\leq i\neq j\leq 3}\{a_0\vec{\eta} + b_0\vec{\nu}_{ij} + c_0\vec{\mu_i} \in \Omega : \frac{a_0}{C_\eta} + \frac{b_0}{C_{\nu_{ij}}} + \frac{c_0}{C_{\mu_i}} \leq \gamma\}.$$

The solution $S(x)$ of (5.26) can be defined in $\Omega(\gamma)$ as follows

$$S(x(0)) = \int_0^T p(t)\, dx(t) \quad \text{for } x(0) \in \Omega(\gamma) \tag{5.54}$$

along the path $(x(t), p(t))$ constructed above.

From (5.54) we have $S(0) = 0$, and it follows by an adaptation of van der Schaft's method [van der Schaft 1996]:

$$H(x, S_x^T(x)) = H(x, p) \equiv 0$$

for any $x \in \Omega(\gamma)$.

Since $x(t)$ is piecewisely smooth as constructed above, we have

$$S(x(0)) = \int_0^T p(t)\, \dot{x}(t) dt \tag{5.55}$$

with the path $(x(t), p(t))$ whose components are of the forms (5.51), (5.52), and (5.53). Plugging these components into (5.55), it can easily be verified that

$$S(x(0)) \geq 0$$

for any $x(0) \in \Omega(\gamma)$.

This $S(x)$ is constructed so that the state-feedback control $x \to u_*(x) = u^*(x, S_x^T(x))$ is of the form:

$$u_*(x) = e_i \quad \text{if } x \in X_i \cup X_{ij} \cup X_{123} \tag{5.56}$$

where $i \neq j$, $i, j \in I$.

Now we may prove the following theorem.

Theorem 5.1 *Let $S(x)$, $x \in \Omega(\gamma)$ be the function constructed as above, i.e., $S(x)$ satisfies*

$$H(x, S_x^T(x)) = 0, \quad S(x) \geq 0, \text{and } S(0) = 0 \tag{5.57}$$

Let the set-valued state-feedback $x \to u_*(x)$ according to the rule

$$u_*(x) = u^*(x, S_x^T(x)).$$

Then the dissipation inequality

$$S(x(t_2)) - S(x(t_1)) \leq \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\, dt \tag{5.58}$$

satisfied along all trajectories $(x(t), q(t))$ of the closed loop system (5.14) (where the solutions are taken in the sense of Filippov) such that the state vector $x(t)$ remains in $\Omega(\gamma)$ over the time interval $[t_1, t_2]$.

In particular, the L_2-gain property

$$\int_0^T \|x(t)\|^2\, dt \leq \gamma^2 \int_0^T \|q(t)\|^2\, dt + S(x(0)) \tag{5.59}$$

holds over all trajectories $(x(t), q(t))$ with $x(t) \in \Omega(\gamma)$ for $0 \leq t \leq T$.

Proof Suppose that $(x(t), q(t))$ is a solution of (5.9), then

$$0 = H(x, S_x^T(x)) = H_{pre}(x, p, u^*)$$

where u^* is any choice of element in $u^*(x, S_x^T(x))$ and $H_{pre}(x, p, u^*)$ defined in (5.16). Thus we have

$$S_x(x)(q(t) - Gu^*(t)) - \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) \leq 0$$

by the first inequality of (5.11), or,

$$\frac{d}{dt}S(x) - \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) \leq 0 \tag{5.60}$$

as long as $x(t)$ remains in the region $\Omega(\gamma)$. Integrating (5.60) from t_1 to t_2, we obtain the dissipation inequality (5.58)

$$S(x(t_2)) - S(x(t_1)) \leq \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\, dt$$

Now we take the time interval $[t_1, t_2]$ as $[0, T]$ and rearrange the above inequality, this becomes

$$S(x(T)) \leq \gamma^2 \int_0^T \|q(t)\|^2\, dt + S(x(0)) - \int_0^T \|x(t)\|^2\, dt$$

By the fact $S(x) \geq 0$, and hence $S(x(T)) \geq 0$. We obtain the L_2-gain property (5.59)

$$\int_0^T \|x(t)\|^2 \, dt \leq \gamma^2 \int_0^T \|q(t)\|^2 \, dt + S(x(0))$$

This completes the proof of the theorem.

In practice, we may adapt the control law as follows: the green time will start from g_{min}, after this initial period, the controller compares the weighted queue lengths $s_i x_i$ $(i \in I)$ from the three approaches, if the current one is still the biggest, then extends the green time by one unit. Then it compares the weighted queue lengths again, until another approach has the biggest weighted queue or is reaching its uplimit of queue, or the current phase is reaching g_{max}. The controller will switch the green into that approach associated with the biggest weighted queue or with the fully loaded capacity, and so on. This leads to the following algorithm:

Algorithm 5.1 *At each instant of time, the controller keeps track of the following variables: the three state variables representing queue lengths x_1, x_2, and x_3, and the time spent in the current state T_c. Let x_c be the queue length of the approach for which the current signal is green. Let $x_o^i \, (i = 1, 2)$ be the queue lengths of the other two approaches, the phase associated with x_o^1 be closer to the current one than the phase associated with x_o^2, and x_{max}^i be the uplimit of x_o^i where $i = 1, 2$. We assume that the minimum green time g_{min}, the unit extension τ, the maximum green time g_{max}, $N = [\frac{g_{max} - g_{min}}{\tau}]$ be the largest integer in the closed interval $[0, \frac{g_{max} - g_{min}}{\tau}]$, the saturation flow rate s_c for the current approach and $s_o^i \, (i = 1, 2)$ for the other two approaches are specified. Then the controller is implemented as follows:*

1. *when $T_c \leq g_{min}$, stay in the current state.*
2. *when $g_{min} + (k-1)\tau < T_c \leq g_{min} + k\tau < g_{max}$, $(k = 1, 2, \cdots, N)$*
   - (a) *switch green into the approach associated with x_o^1 at the end of this interval if $x_o^1 \geq x_{max}^1$,*
   - (b) *switch green into the approach associated with x_o^2 at the end of this interval if $x_o^2 \geq x_{max}^2$,*
   - (c) *switch green into the approach associated with x_o^1 at the end of this interval if $s_c x_c < s_o^1 x_o^1$ and $s_o^1 x_o^1 \geq s_o^2 x_o^2$,*
   - (d) *switch green into the approach associated with x_o^2 at the end of this interval if $s_c x_c < s_o^2 x_o^2$ and $s_o^1 x_o^1 < s_o^2 x_o^2$, or*
   - (e) *stay in the current state if $s_c x_c \geq \max\{s_o^1 x_o^1, s_o^2 x_o^2\}$ and $x_o^i < x_{max}^i$ for both $i = 1, 2$.*
3. *when $T_c \geq g_{max}$, switch green to the approach associated with x_o^1 if $s_o^1 x_o^1 \geq s_o^2 x_o^2$, or to the approach associated with x_o^2 if $s_o^1 x_o^1 < s_o^2 x_o^2$.*

5.3 A multi-phase intersection

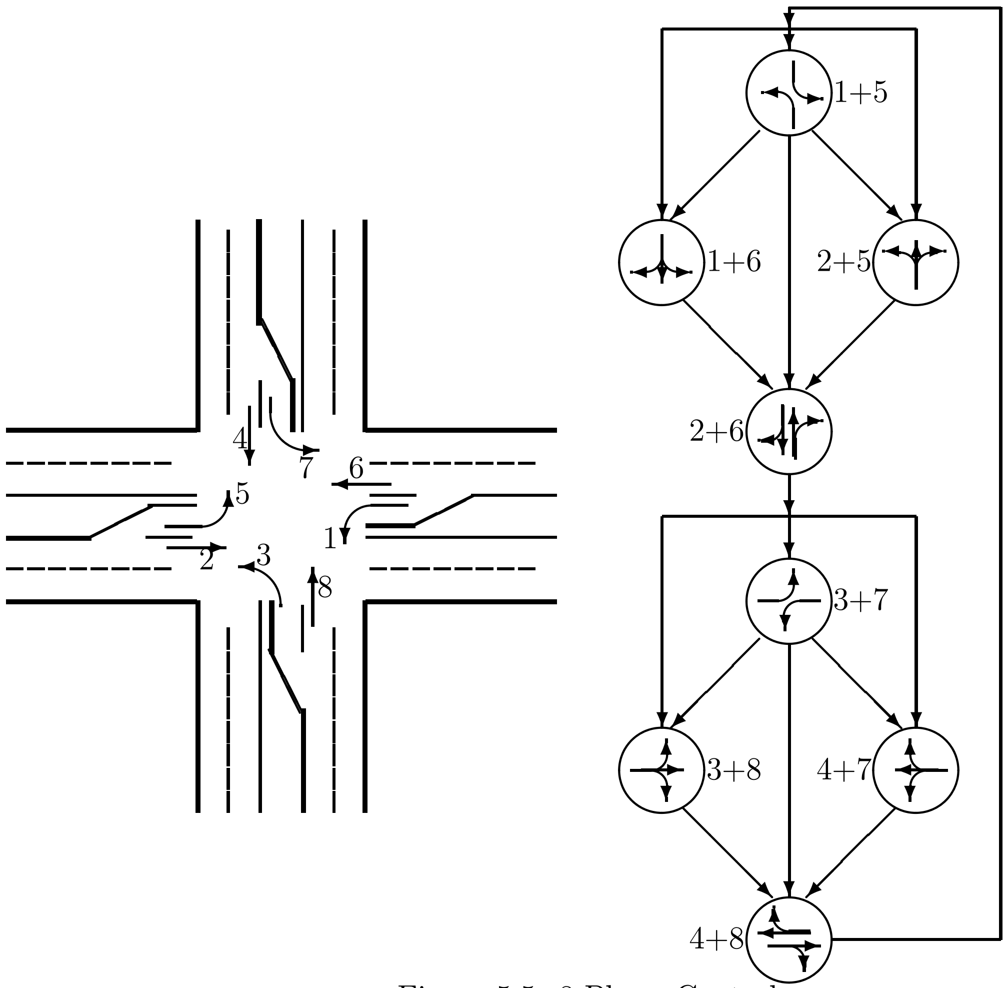

Figure 5.5: 8-Phase Control

A more complicated example is given in Figure 5.5. Here the traffic signal has 8 phases. Now Let us consider a general n-phase intersection and denote $I_n = \{1, 2, \cdots, n\}$.

State Variables:

x_i: the queue length of the traffic stream in phase $\phi_i\,(i \in I_n)$

Exogenous inputs:

q_i: the arrival rate of the vehicles toward phase $\phi_i\,(i \in I_n)$

Parameters:

s_i: the saturation flow rate of phase $\phi_i\,(i \in I_n)$

Control Variables:

Let $\mathcal{U} = \{e_1, e_2, \ldots, e_n\}$ be a standard base of $\mathbf{R}^n$, $u \in \mathcal{U}$: when $u = e_i\,(i \in I_n)$, phase ϕ_i is green.

We assume all s_i to be strictly positive where $i \in I_n$. All other quantities are assumed or constrained to be nonnegative.

We denote $conv\,\mathcal{U}$ as the convex hull of $\mathcal{U}$, i.e.,

$$conv\,\mathcal{U} = \{\lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n : \sum_{i=0}^{n} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0 \quad i = 0, 1, \cdots, n\}.$$

and let Ω be a connected region of the interior of

$$\mathbf{R}^{n+} = \{x = (x_1, x_2, \cdots, x_n)^T : x_i \geq 0,\, i \in I_n\}.$$

Similarly we set up the state dynamic equations:

$$\begin{cases} \dot{x}_1 &= \delta(x_1, q_1 - [s_1, 0, \ldots, 0]u) \\ \dot{x}_2 &= \delta(x_2, q_2 - [0, s_2, \ldots, 0]u) \\ &\vdots \\ \dot{x}_n &= \delta(x_n, q_n - [0, 0, \ldots, s_n]u) \end{cases} \tag{5.61}$$

where $\delta(\cdot,\cdot)$ is defined in (4.5). In the oversaturated case this simplifies to

$$\begin{cases} \dot{x}_1 &= q_1 - [s_1, 0, \ldots, 0]u) \\ \dot{x}_2 &= q_2 - [0, s_2, \ldots, 0]u) \\ &\vdots \\ \dot{x}_3 &= q_n - [0, 0, \ldots, s_n]u) \end{cases} \tag{5.62}$$

Similarly we have the projection dynamics form of the model (5.61)

$$\dot{x} = \pi_K(x, q - Gu)), \quad x(0) = x_0 \in \Omega. \tag{5.63}$$

where G is a diagonal matrix of the form $\begin{pmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n \end{pmatrix}$.

The pre-Hamiltonian is

$$K_\gamma(x, p, q, u) = p^T(q - Gu) - \frac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2) \tag{5.64}$$

As the previous sections we have the critical q given by $q^*(x,p) = \frac{1}{\gamma^2}p$ and the semi-pre-Hamiltonian $H_{pre,\gamma}(x,p,u) = \max_{q \in \mathbf{R}^n} K_\gamma(x,p,q,u)$ is given by

$$H_{pre,\gamma}(x, p, u) = \frac{1}{2\gamma^2}\|p\|^2 - p^T Gu + \frac{1}{2}\|x\|^2 \tag{5.65}$$

Let us assign Hamiltonian $H_{1,\gamma}(x,p) = H_{pre,\gamma}(x,p,e_1)$, $H_{2,\gamma}(x,p) = H_{pre,\gamma}(x,p,e_2), \cdots$, and $H_{n,\gamma}(x,p) = H_{pre,\gamma}(x,p,e_n)$. Explicitly, we have

$$H_{1,\gamma}(x,p) = \frac{1}{2\gamma^2}\|p\|^2 - s_1 p_1 + \frac{1}{2}\|x\|^2 \tag{5.66}$$

$$H_{2,\gamma}(x,p) = \frac{1}{2\gamma^2}\|p\|^2 - s_2 p_2 + \frac{1}{2}\|x\|^2. \tag{5.67}$$

$$\vdots$$

$$H_{n,\gamma}(x,p) = \frac{1}{2\gamma^2}\|p\|^2 - s_n p_n + \frac{1}{2}\|x\|^2. \tag{5.68}$$

The Hamiltonian $H_\gamma(x,p) = \min_{u\in\mathcal{U}} H_{pre,\gamma}(x,p,u)$ is given simply by

$$H_\gamma(x,p) = \min\{H_{1,\gamma}(x,p),\ H_{2,\gamma}(x,p), \cdots, H_{n,\gamma}(x,p)\}. \tag{5.69}$$

We may complete the squares in $H_{i,\gamma}(x,p)\,(i \in I_n)$ to obtain the alternate expressions

$$H_{1,\gamma}(x,p) = \frac{1}{2\gamma^2}(p_1 - \gamma^2 s_1)^2 + \frac{1}{2\gamma^2}(p_2^2 + p_3^2 + \cdots + p_n^2) - \frac{\gamma^2}{2}s_1^2 + \frac{1}{2}\|x\|^2, \tag{5.70}$$

$$H_{2,\gamma}(x,p) = \frac{1}{2\gamma^2}(p_1^2 + p_3^2 + \cdots + p_n^2) + \frac{1}{2\gamma^2}(p_2 - \gamma^2 s_2)^2 - \frac{\gamma^2}{2}s_2^2 + \frac{1}{2}\|x\|^2 \tag{5.71}$$

$$\vdots$$

$$H_{n,\gamma}(x,p) = \frac{1}{2\gamma^2}(p_1^2 + p_2^2 + \cdots + p_{n-1}^2) + \frac{1}{2\gamma^2}(p_n - \gamma^2 s_n)^2 - \frac{\gamma^2}{2}s_n^2 + \frac{1}{2}\|x\|^2 \tag{5.72}$$

We define the following regions in the space of costates:

$$\Pi_i = \{(p_1, p_2, \cdots, p_n)^T : p_i > 0 \text{ and } s_i p_i > \max_{1\le j\ne i\le n}\{s_j p_j\}\}$$

and the corresponding regions in the state space:

$$X_i = \{x \in \Omega : s_i x_i > \max_{1\le j\ne i\le n}\{s_j x_j\}\}$$

Similarly to the above section where $n = 3$, we can construct the storage function $S(x)$ satisfying

$$H(x, S_x^T(x)) = 0, \quad S(x) \ge 0, \text{and } S(0) = 0 \tag{5.73}$$

when x in an appropriate region $\Omega(\gamma) \subset \Omega$.

The following theorem holds true for the n-phase intersection.

Theorem 5.2 *Let $S(x)$, $x \in \Omega(\gamma)$ be the function satisfying (5.73), and define the set-valued state-feedback $x \to u_*(x)$ according to the rule*

$$u_*(x) = u^*(x, S_x^T(x))$$

Then the dissipation inequality

$$S(x(t_2)) - S(x(t_1)) \leq \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2 \|q(t)\|^2 - \|x(t)\|^2)\, dt \tag{5.74}$$

is satisfied along all trajectories $(x(t), q(t))$ of the following closed loop system:

$$\dot{x}(t) = \pi_K(x, q - Gu_*(x)) \tag{5.75}$$

(where the solutions are taken in the sense of Filippov) such that the state vector $x(t)$ remains in $\Omega(\gamma)$ over the time interval $[t_1, t_2]$.

In particular, the L_2-gain property

$$\int_0^T \|x(t)\|^2\, dt \leq \gamma^2 \int_0^T \|q(t)\|^2\, dt + S(x(0)) \tag{5.76}$$

holds over all trajectories $(x(t), q(t)$ with $x(t) \in \Omega(\gamma)$ for $0 \leq t \leq T$.

Full details of this n-phase signal control will appear in [Ball, Day and Kachroo 1997].

5.4 Conclusion

For the general n-phase intersection, we have the traffic-actuated control as follows: the green time will start from g_{min}, after this initial period, the controller compares the weighted queue lengths $s_i x_i$ ($i \in I_n$) from all n approaches, if the current one is still the biggest, then extends the green time by one unit. Then it compares the weighted queue lengths again, until another approach has the biggest weighted queue or is reaching its uplimit of queue, or the current phase is reaching g_{max}. The controller will switch the green into that approach associated with the biggest weighted queue or with the fully loaded capacity, and so on.

Algorithm 5.2 *At each instant of time, the controller keeps track of the following variables: the state variables representing queue lengths x_i ($i \in I_n$), and the time spent in the current state T_c. Let x_c be the queue length of the approach for which the current signal is green. Let x_o^i ($i = 1, 2, \cdots, n-1$) be the queue lengths of the other ($n-1$) approaches, and x_{max}^i be the uplimit of x_o^i ($i = 1, 2, \cdots, n-1$). We assume that the minimum green time g_{min}, the unit extension τ, the maximum green time g_{max}, $N = [\frac{g_{max}-g_{min}}{\tau}]$ be the largest integer in the closed interval $[0, \frac{g_{max}-g_{min}}{\tau}]$, the saturation flow rate s_c for the current approach and s_o^i ($i = 1, 2, \cdots, n-1$) for the other ($n-1$) approaches are specified. Then the traffic-actuated controller is implemented as follows:*

1. *when $T_c \leq g_{min}$, stay in the current state.*

2. *when* $g_{min} + (k-1)\tau < T_c \leq g_{min} + k\tau < g_{max}, \quad (k = 1, 2, \cdots, N)$
   (a) *switch green into the approach associated with* x_o^j *at the end of this interval if* $x_o^j \geq x_{max}^j$ *for* $i = 1, 2, \cdots, n-1$
   (b) *switch green into the approach associated with* x_o^j *at the end of this interval if* $s_c x_c < s_o^j x_o^j$ *and* $s_o^j x_o^j \geq \max\{s_o^i x_o^i : i = 1, 2, \cdots, n-1\}$*, or*
   (c) *stay in the current state if* $s_c x_c \geq \max\{s_o^i x_o^i : i = 1, 2, \cdots, n-1\}$ *and* $x_o^i < x_{max}^i$ *for all* $i = 1, 2, \cdots, n-1$.
3. *when* $T_c \geq g_{max}$, *switch green to the approach associated with* x_o^j *if* $s_o^j x_o^j \geq \max\{s_o^i x_o^i : i = 1, 2, \cdots, n-1\}$.

In case there are more than one approach satisfied the same condition, the controller will always pick the closest phase to the current one.

Chapter 6

Traffic Network Signalization

This chapter attempts to develop new traffic models for arterial street control. In contrast with isolated intersections, the signals must operate as a interactive system. Arterial street control recognizes that a signal releases platoons that travel to the next signal. To maintain the flow of these platoons, the system must coordinate timing of adjacent intersections. The system accomplishes this by establishing a time relationship between the beginning of arterial green at one intersection and the beginning of arterial green at the next intersection. This permits continuous traffic flow along an arterial street and reduces delay.

6.1 A model of an arterial network

We model an arterial network as an ordered graph. The nodes of the graph correspond to detector stations. Let $I \subset \mathbf{Z}^+$ denote the index set of the nodes. At a node there may be an entry, exit, intersection, or only the joint of two sections. Let $I_O \subset I$ denote the index set of the nodes at which there are entries or origins of the system. Let $I_D \subset I$ denote the index set of the nodes at which there are exits or destinations of the system. $I_I \subset I$ denote the index set of the nodes at which there are intersections. A *section* of the system is an ordered pair $(l, m) \in I \times I$ such that there is a path from node l to node m not passing another node. Let $S \subset I \times I$ denote the set of sections. An *origin-destination pair*, usually called OD-pair, is an ordered tuple $(i, j) \in I_O \times I_D$, where $i \in I_O$ is a node with an entry or origin and $j \in I_D$ is a node with an exit or destination. Let $OD \subset I_O \times I_D$ be the set of OD-pairs. A *route* of the OD-pair (i, j) is a path from node i to node j. A route can be specified by a chain of sections. A *link* is a chain of sections between two adjacent intersections of the system. In general an OD-pair may admit several routes. For $(i, j) \in OD$ let $K(i, j) \subset \mathbf{Z}^+$ be the index set of distinct possible routes and let $R(i, j)$ denote the corresponding set of routes, that is,

$$R(i, j) = r((i, j), k), k \in K(i, j)$$

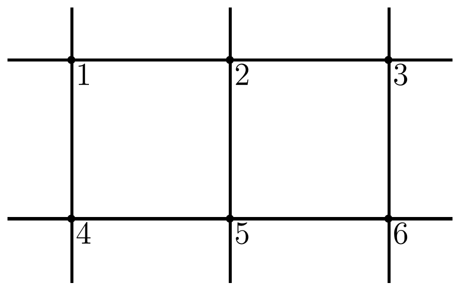

Figure 6.1: A Sample Closed Network

A route is described as a chain of sections

$$r((i,j),k) = (i,i_2),(i_3,i_4),...,(i_s,j).$$

According to Payne's model [Payne 1971][Van Schuppen 1992], the state variables of traffic flow are density and average speed of each section.

Denote the density of OD-pair $(i,j) \in OD$, of route number $k \in K(i,j)$, and of section $(l,m) \in S$ by $\rho((i,j),k,(l,m),\cdot) : T \rightarrow \mathbf{R}^+$ in vehicles per km per lane. Denote the density in section $(l,m) \in S$ by $\rho((l,m),\cdot) : T \rightarrow \mathbf{R}^+$

$$\rho((l,m),t) = \sum \rho((i,j),k,(l,m),t)$$

where the sum is over all $(i,j) \in OD$ and all routes $k \in K(i,j)$ that use section (l,m).

Denote the average speed of all vehicles in section (l,m) by $v((l,m),\cdot) : T \rightarrow \mathbf{R}^+$. Thus, the flow of section (l,m) can be defined as $q((l,m),\cdot) : T \rightarrow \mathbf{R}^+$ in vehicles per hour.

$$q((l,m),t) = N(l,m) \times \rho((l,m),t) \times v((l,m),t)$$

where $N(l,m)$ is the number of lanes of section (l,m).

Where two arterials cross at an intersection, a signal timing interlock must occur for progression along both arterials. Signal timing in closed networks conventionally features a common cycle length. The closed topology of the network generally requires a constraint on the offsets. The sum of offsets around each loop in the network must have sum equal to an integral number of cycle lengths. For example, we have the following relationships for the sample closed network in Figure 6.1:

$$\begin{cases} \Theta_{1,2} + \Theta_{2,3} + \Theta_{3,6} + \Theta_{6,5} + \Theta_{5,4} + \Theta_{4,1} = m\,C \\ \Theta_{1,2} + \Theta_{2,5} + \Theta_{5,4} + \Theta_{4,1} = n\,C \\ \Theta_{2,3} + \Theta_{3,6} + \Theta_{6,5} + \Theta_{5,2} = (m-n)\,C \end{cases} \tag{6.1}$$

where $\Theta_{i,j}$ is the offset between nodes i and j, C is the common cycle length, m and n are positive integers.

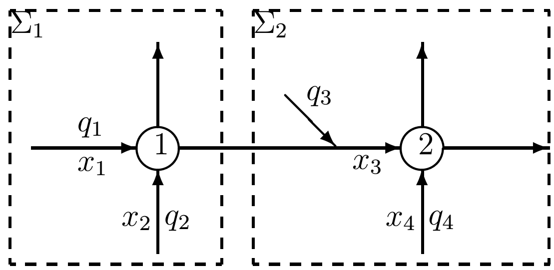

Figure 6.2: A Simple Arterial Street (Σ) with Two Intersections

6.2 A simple arterial street with two intersections

A simple arterial street Σ with two intersections is shown in Figure 6.2. The problem is formulated by using the following notation:

State Variables:
x_i: the queue length of the traffic stream on approach i, $\quad i = 1, 2, 3, 4$.

Exogenous inputs:
q_i: the arrival flow rate on approach i, $\quad i = 1, 2, 3, 4$.

Parameters:
s_i: the saturation flow rate on approach i, $\quad i = 1, 2, 3, 4$.

Control Variables:
$\mathbf{U} \equiv \{u^1, u^2, u^3, u^4\}$: when $u = u^1$, approaches 1 and 3 are green; when $u = u^2$, approaches 1 and 4 are green; when $u = u^3$, approaches 2 and 3 are green; and when $u = u^4$, approaches 2 and 4 are green; We may normalize the four possible control settings to be the four standard basis vectors $u^1 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^T$, $u^2 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}^T$, $u^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}^T$, $u^4 = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^T$ in $\mathbf{R}^4$,

We assume all links are single lanes, all vehicles have no turns but go straightly downstream, and all $s_i (i = 1, 2, 3, 4)$ to be strictly positive. All other quantities are assumed or constrained to be nonnegative.

Various models are possible, depending on the underlying assumptions. However, from the theory of dissipative systems, the interconnected system Σ is expected to be dissipative, since both Σ_1 and Σ_2 are dissipative.

6.3 Cascade connection without delay

Assume that the flow served at intersection 1 is delivered instantaneously to intersection 2 to add to the queue there. Following steps similar to those in section 4.1, we arrive at the state dynamic equations

$$\begin{cases} \dot{x}_1 &= \delta(x_1, q_1 + (-s_1 \quad -s_1 \quad 0 \quad 0)u) \\ \dot{x}_2 &= \delta(x_2, q_2 + (0 \quad 0 \quad -s_2 \quad -s_2)u) \\ \dot{x}_3 &= \delta(x_3, q_3 + (s_1 \ \mathrm{sgn}(x_1) - s_3 \quad s_1 \ \mathrm{sgn}(x_1) \quad -s_3 \quad 0)u) \\ \dot{x}_4 &= \delta(x_4, q_4 + (0 \quad -s_4 \quad 0 \quad -s_4)u) \end{cases} \tag{6.2}$$

where δ is defined in (4.5) and

$$\mathrm{sgn}(y) = \begin{cases} 1, & y > 0 \\ 0, & y = 0 \end{cases}$$

for a variable y constrained to satisfy $y \geq 0$.

The system (6.2) can be written in vector notation to

$$\dot{x} = \delta(x, q - G(x, u)) \tag{6.3}$$

where $x = (x_1, x_2, x_3, x_4)^T$, $q = (q_1, q_2, q_3, q_4)^T$, $\delta(x, y) = (\delta(x_1, y_1), \delta(x_2, y_2), \delta(x_3, y_3), \delta(x_4, y_4))^T$, and

$$G(x, u) = \begin{pmatrix} s_1 & s_1 & 0 & 0 \\ 0 & 0 & s_2 & s_2 \\ s_3 - s_1 \ \mathrm{sgn}(x_1) & -s_1 \ \mathrm{sgn}(x_1) & s_3 & 0 \\ 0 & s_4 & 0 & s_4 \end{pmatrix} u.$$

We didn't model the link capacity in (6.2), the model will not work when the demand of the link between two intersections exceeds the maximum capacity of the link. That is, when the queue in approach 3 spills back to intersection 1, the model (6.2) is no longer valid and we have a jam condition. In this situation, the flow through intersection 1 cannot be served, even in the presence of a green signal. Practically the green phase should turn to favor approach 3 and approach 2. This scenario leads to the more complicated model:

$$\begin{cases} \dot{x}_1 &= \delta(x_1, q_1 + (\ -s_1 \quad -s_1 \quad 0 \quad 0\ )u \ \ \mathrm{sgn}(\ell - x_3)) \\ \dot{x}_2 &= \delta(x_2, q_2 + (\ 0 \quad 0 \quad -s_2 \quad -s_2\ )u) \\ \dot{x}_3 &= \delta(x_3, q_3 + (\ s_1 \ \mathrm{sgn}(x_1) - s_3 \quad s_1 \ \mathrm{sgn}(x_1) \quad -s_3 \quad 0\ )u) \\ \dot{x}_4 &= \delta(x_4, q_4 + (\ 0 \quad -s_4 \quad 0 \quad -s_4\ )u \ \ \mathrm{sgn}(\ell - x_3)) \end{cases} \tag{6.4}$$

where ℓ is the maximum queue length allowed in the link between two intersections.

We have the similar matrix format to (6.3) but here

$$G(x, u) = \begin{pmatrix} s_1 \ \mathrm{sgn}(\ell - x_3) & s_1 \ \mathrm{sgn}(\ell - x_3) & 0 & 0 \\ 0 & 0 & s_2 & s_2 \\ s_3 - s_1 \ \mathrm{sgn}(x_1) & -s_1 \ \mathrm{sgn}(x_1) & s_3 & 0 \\ 0 & s_4 \ \mathrm{sgn}(\ell - x_3) & 0 & s_4 \ \mathrm{sgn}(\ell - x_3) \end{pmatrix} u.$$

Remark 1: The models (6.2) and (6.4) may not be feasible for the traffic network, but they are useful models for some queueing networks where the transition time between nodes can be ignored compared to the total delay on the network.

Theorem 6.1 *The interconnected system* Σ *in (6.2) is dissipative*

Proof: For simplicity, we take $\gamma = 1$.

According to chapter 4, we know that both Σ_1 and Σ_2 are dissipative, that is, there are two storage functions S_1 and S_2 such that

$$S_1(x(t_2)) - S_1(x(t_1)) \leq \frac{1}{2}\int_{t_1}^{t_2}(\|q_1\|^2 + \|q_2\|^2 - \|q_o\|^2 - \|z_2\|^2)dt \tag{6.5}$$

where q_o is the output from Σ_1 to Σ_2 and z_2 is the other output from Σ_1 in approach 2.

$$S_2(x(t_2)) - S_2(x(t_1)) \leq \frac{1}{2}\int_{t_1}^{t_2}(\|q_o\|^2 + \|q_3\|^2 + \|q_4\|^2 - \|z_3\|^2 - \|z_4\|^2)dt \tag{6.6}$$

where z_3 and z_4 are the outputs from Σ_2 in approach 3 and 4, respectively.

Adding (6.5) and (6.6) together, we have

$$\begin{aligned} S_1(x(t_2)) + S_2(x(t_2)) - S_1(x(t_1)) - S_2(x(t_1)) \quad \leq \quad & \tfrac{1}{2}\textstyle\int_{t_1}^{t_2}(\|q_1\|^2 + \|q_2\|^2 + \|q_3\|^2 + \|q_4\|^2 \\ & -\|z_2\|^2 - \|z_3\|^2 - \|z_4\|^2)dt \end{aligned}$$

Setting $S(x) = S_1(x) + S_2(x)$, we obtain

$$S(x(t_2)) - S(x(t_1)) \leq \frac{1}{2}\int_{t_1}^{t_2}(\|q_1\|^2 + \|q_2\|^2 + \|q_3\|^2 + \|q_4\|^2 - \|z_2\|^2 - \|z_3\|^2 - \|z_4\|^2)dt$$

That is,

$$S(x(t_2)) - S(x(t_1)) \leq \frac{1}{2}\int_{t_1}^{t_2}(\|q\|^2 - \|z\|^2)dt$$

where $z = (z_2, z_3, z_4)^T$.

Therefore, Σ is dissipative with respect to the supply rate $s(q, z) = \|q\|^2 - \|z\|^2$.

This completes the proof.

6.4 Cascade connection with delay

Assume that the average travel time between intersections 1 and 2 is τ. Then the rate of increase of the queue length at intersection 2 attributable to the release of traffic at intersection 1 is actually equal to the flow served at intersection 1 at τ units of time earlier. If we ignore the possibility of gridlock, this scenario leads to the following dynamical equations:

$$\left\{ \begin{array}{lcl} \dot{x}_1 & = & \delta(x_1, q_1 + (\ -s_1 \quad -s_1 \quad 0 \quad 0\ )u) \\ \dot{x}_2 & = & \delta(x_2, q_2 + (\ 0 \quad 0 \quad -s_2 \quad -s_2\ )u) \\ \dot{x}_3 & = & \delta(x_3, q_3 + s_1\ \operatorname{sgn}(x_1)\, u_1(\cdot - \tau) - s_3 u_1 + s_1\ \operatorname{sgn}(x_1)\, u_2(\cdot - \tau) - s_3 u_3) \\ \dot{x}_4 & = & \delta(x_4, q_4 + (\ 0 \quad -s_4 \quad 0 \quad -s_4\ )u) \end{array} \right. \tag{6.7}$$

where u_1, u_2, u_3, u_4 represent the components of the vector $u = (\ u_1 \quad u_2 \quad u_3 \quad u_4\ )^T$ at a given time t. The presence of the time delay τ in parts of the control vector in these equations makes this case difficult to handle mathematically.

The system (6.7) can be written in vector notation to

$$\dot{x} = \delta(x, q - G(x, u)) \tag{6.8}$$

where x, q, and δ are defined as in (6.2), but

$$G(x,u) = \begin{pmatrix} s_1 u_1 & s_1 u_2 & 0 & 0 \\ 0 & 0 & s_2 u_3 & s_2 u_4 \\ s_3 u_1 - s_1\ \operatorname{sgn}(x_1) u_1(\cdot - \tau) & -s_1\ \operatorname{sgn}(x_1) u_2(\cdot - \tau) & s_3 u_3 & 0 \\ 0 & s_4 u_2 & 0 & s_4 u_4 \end{pmatrix}.$$

If we consider the maximum queue limit (ℓ) in the middle link between two intersections, we have the following model:

$$\left\{ \begin{array}{lcl} \dot{x}_1 & = & \delta(x_1, q_1 + (\ -s_1 \quad -s_1 \quad 0 \quad 0\ )u\ \operatorname{sgn}(\ell - x_3)) \\ \dot{x}_2 & = & \delta(x_2, q_2 + (\ 0 \quad 0 \quad -s_2 \quad -s_2\ )u) \\ \dot{x}_3 & = & \delta(x_3, q_3 + s_1\ \operatorname{sgn}(x_1)\, u_1(\cdot - \tau) - s_3 u_1 + s_1\ \operatorname{sgn}(x_1)\, u_2(\cdot - \tau) - s_3 u_3) \\ \dot{x}_4 & = & \delta(x_4, q_4 + (\ 0 \quad -s_4 \quad 0 \quad -s_4\ )u\ \operatorname{sgn}(\ell - x_3)) \end{array} \right. \tag{6.9}$$

and thus the corresponding $G(x, u)$ for the matrix format (6.8) is

$$G(x,u) = \begin{pmatrix} s_1 u_1\ \operatorname{sgn}(\ell - x_3) & s_1 u_2\ \operatorname{sgn}(\ell - x_3) & 0 & 0 \\ 0 & 0 & s_2 u_3 & s_2 u_4 \\ s_3 u_1 - s_1\ \operatorname{sgn}(x_1) u_1(\cdot - \tau) & -s_1\ \operatorname{sgn}(x_1) u_2(\cdot - \tau) & s_3 u_3 & 0 \\ 0 & s_4 u_2\ \operatorname{sgn}(\ell - x_3) & 0 & s_4 u_4\ \operatorname{sgn}(\ell - x_3) \end{pmatrix}.$$

6.5 Networking with link dynamics.

We introduce another state variable ρ as the density of the middle link between two intersections. We assume a maximum average velocity v_f and a maximum density ρ_{max}. Then we may use the simple average velocity model proposed by Greenshield [McShane and Roess 1990],

$$v = v_f(1 - \frac{\rho}{\rho_{max}})$$

.

Assumption 6.1 *The dynamic rate of the link density represents exactly the difference between the incoming vehicles per km per hour and the exiting vehicles per km per hour at the link.*

This Assumption leads to the differential equation

$$\dot{\rho} = \begin{pmatrix} s_1 & s_1 & 0 & 0 \end{pmatrix} u \;\; \mathrm{sgn}(x_1(\rho_{max} - \rho)) - v_f(1 - \frac{\rho}{\rho_{max}})\rho.$$

Since ρ_{max} implies the maximum queue length allowed in the middle link, it is natural to model the link capacity in this case. Therefore, we develop the following network model with link dynamics:

$$\left\{ \begin{array}{lcl} \dot{x}_1 & = & \delta(x_1, q_1 + \begin{pmatrix} -s_1 & -s_1 & 0 & 0 \end{pmatrix} u \; \mathrm{sgn}(\rho_{max} - \rho)) \\ \dot{x}_2 & = & \delta(x_2, q_2 + \begin{pmatrix} 0 & 0 & -s_2 & -s_2 \end{pmatrix} u) \\ \dot{\rho} & = & \begin{pmatrix} s_1 & s_1 & 0 & 0 \end{pmatrix} u \;\; \mathrm{sgn}(x_1(\rho_{max} - \rho)) - v_f(1 - \frac{\rho}{\rho_{max}})\rho \\ \dot{x}_3 & = & \delta(x_3, q_3 + v_f(1 - \frac{\rho}{\rho_{max}})\rho + \begin{pmatrix} -s_3 & 0 & -s_3 & 0 \end{pmatrix} u) \\ \dot{x}_4 & = & \delta(x_4, q_4 + \begin{pmatrix} 0 & -s_4 & 0 & -s_4 \end{pmatrix} u \, \mathrm{sgn}(\rho_{max} - \rho)) \end{array} \right. . \tag{6.10}$$

In the oversaturated case the system (6.10) can be written in vector notation to

$$\dot{x} = q - G(x, u) \tag{6.11}$$

where $x = (x_1, x_2, \rho, x_3, x_4)^T$, $q = (q_1, q_2, 0, q_3, q_4)^T$, $u = (u_1, u_2, u_3, u_4)^T$, and

$$\begin{array}{rcl} G(x,u) & = & \begin{pmatrix} s_1 \mathrm{sgn}(\rho_{max} - \rho) & s_1 \mathrm{sgn}(\rho_{max} - \rho) & 0 & 0 \\ 0 & 0 & s_2 & s_2 \\ -s_1 \, \mathrm{sgn}(x_1(\rho_{max} - \rho)) & -s_1 \, \mathrm{sgn}(x_1(\rho_{max} - \rho)) & 0 & 0 \\ s_3 & 0 & s_3 & 0 \\ 0 & s_4 \, \mathrm{sgn}(\rho_{max} - \rho) & 0 & s_4 \, \mathrm{sgn}(\rho_{max} - \rho) \end{pmatrix} u \\ & & + \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} v_f(1 - \frac{\rho}{\rho_{max}})\rho. \end{array}$$

6.6 Some results for a class of queueing networks

Queues are a common aspect of any irregular process operating in an environment with limited capacity. A **queueing network** consists of a service facility that provides service of some kind to arriving customers. The service facility requires a certain time to serve each customer and is capable of serving only a limited number of customers at a time. If customers arrive faster than the facility can serve them, they must wait in a queue for the facility. In the traffic network, customers are vehicles and servers are the transportation infrastructure such as links and intersections. In the computer network, customers are data and servers are the internet infrastructure such as routes, local-area networks, wide-area networks, network file servers, and computers. *Congestion* will occur in the system if there is sufficient irregularity.

We may set up the state dynamic equations for a class of queueing networks satisfying the following assumption:

Assumption 6.2 *The dynamic rate of the queue length at each node represents exactly the difference between the arrival rate and the service rate at that node, as long as either the queue length to be served is positive or the arrival rate is greater than the service rate; Otherwise, there is no queue dynamics for the node*

Now we can extend the above traffic network models to this class of the queueing network with the matrix form:

$$\dot{x} = \delta(x, q - G(x, u)) \tag{6.12}$$

where $x(t)$ and $q(t)$ take values in the first orthant $\mathbf{R}^{n+}$, $u(t)$ takes values in the admissible control set $\mathcal{U}$ which is assumed to be the finite set $\mathcal{U} = \{e_i : i = 1, \ldots, n\}$ consisting of the standard basis vectors e_i $(i = 1, 2, \ldots, n)$ in $\mathbf{R}^n$. $G(x, u)$ is continuous in u and measurable in x with a typical form $G(x, u) = g(x)u + f(x)$. The vector-valued function $\delta(x, y) = (\delta(x_1, y_1), \delta(x_2, y_2), \cdots, \delta(x_n, y_n))^T$.

The L_2-gain control problem of the queueing network (6.12) can be defined as

Definition 6.1 (State-feedback L_2-gain problem of the queueing network) *(6.12): Given a tolerance level $\gamma \geq 0$, to design a state-feedback control $x \rightarrow u_*(x)$ so that the resulting closed-loop system*

$$\dot{x}(t) = \delta(x(t), q(t) - G(x(t)), u_*(x(t)), \quad x(0) = x_0 \tag{6.13}$$

satisfies the L_2-gain condition

$$\int_0^T \|x(t)\|^2 \, dt \leq \gamma^2 \int_0^T \|q(t)\|^2 \, dt + \alpha(x_0)$$

for a nonnegative function $x \rightarrow \alpha(x)$ with $\alpha(0) = 0$.

We denote *conv* $\mathcal{U}$ as the convex hull of $\mathcal{U}$, and let Ω be a connected region of the interior of $\mathbf{R}^{n+}$. For simplicity, we consider the oversaturated case with the dynamics:

$$\dot{x} = q - g(x)u - f(x) \tag{6.14}$$

where $g(x)$ is a measurable function and $f(x)$ is a continuous function. So the solutions of this differential inclusion (6.14) is in the sense of Filippov.

We define Hamiltonian function $H(x,p)$ as

$$H(x,p) = \sup_q \inf_{u\in\mathcal{U}} \{p^T(q - g(x)u - f(x)) - \frac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2\}, \tag{6.15}$$

and $H_i(x,p)$ as

$$\begin{aligned} H_i(x,p) = & \sup_q \{p^T(q - g(x)e_i - f(x)) - \tfrac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2)\} & (6.16) \\ = & \tfrac{1}{2\gamma^2}\|p\|^2 - p^T g(x)e_i - p^T f(x) + \tfrac{1}{2}\|x\|^2. & (6.17) \end{aligned}$$

Then, we have

$$H(x,p) = \min_{i=1,2,\ldots,n} H_i(x,p).$$

Note that the minimum over $u \in \mathcal{U}$ and the minimum over $u \in \operatorname{conv}\mathcal{U}$ of the pre-Hamiltonian

$$H_{pre}(x,p,u) := \sup_q \{p^T(q - g(x)u - f(x)) - \frac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2\} \tag{6.18}$$

are the same since $H_{pre}(x,p,u)$ is linear in u.

Theorem 6.2 *Consider the L_2-gain problem (6.1). Let $H(x,p)$ be the Hamiltonian function as defined in (6.15) and $S(x)$ be a piecewise smooth real-valued function satisfying*

$$H(x, S_x^T(x)) = 0, \quad S(x) \geq 0, \quad S(0) = 0 \tag{6.19}$$

for all $x \in \Omega$.

Define $u^*(x,p)$ to be the set

$$u^*(x,p) = \{u \in conv\,\mathcal{U} \colon H_{pre}(x,p,u) = H(x,p)\},$$

and define the set-valued state-feedback $x \to u_*(x)$ according to the rule

$$u_*(x) = u^*(x, S_x^T(x))$$

Then the dissipation inequality

$$S(x(t_2)) - S(x(t_1)) \leq \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\,dt \tag{6.20}$$

is satisfied along all trajectories $(x(t), q(t))$ of the closed loop system (6.13) (where the solutions are taken in the sense of Filippov) such that the state vector $x(t)$ remains in Ω over the time interval $[t_1, t_2]$.

In particular, the L_2-gain property

$$\int_0^T \|x(t)\|^2\,dt \le \gamma^2 \int_0^T \|q(t)\|^2\,dt + S(x(0)) \tag{6.21}$$

holds over all trajectories $(x(t), q(t))$ with $x(t) \in \Omega$ for $0 \le t \le T$.

Proof Suppose that $S(x)$ is a piecewise smooth solution of the Hamilton-Jacobi equation (6.19) with the side conditions $S(0) = 0$ and $S(x) \ge 0$ on the region Ω, and suppose that $(x(t), q(t))$ is a solution of (6.13), then

$$0 = H(x, S_x^T(x)) = H_{pre}(x, S_x^T(x), u^*) \tag{6.22}$$

where u^* is any choice of element in $u^*(x, S_x^T(x))$.

Using $p = S_x^T(x)$ and $u = u^*$ in (6.18) we then have

$$S_x(x)(q - g(x)u^* - f(x)) - \frac{1}{2}(\gamma^2\|q\|^2 - \|x\|^2) \le H_{pre}(x, S_x^T(x), u^*) = 0 \tag{6.23}$$

for all q and u^*.

In particular, take $u^*(t)$ as the element in the set $u^*(x, S_x^T(x))$ associated with the solution $(x(t), q(t))$, then an application of (6.23) gives

$$S_x(x)(q(t) - g(x(t))u^*(t) - f(x(t))) - \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) \le 0$$

or,

$$\frac{d}{dt}S(x(t)) - \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2) \le 0 \tag{6.24}$$

as long as $x(t)$ remains in the region Ω. Integrating (6.24) from t_1 to t_2, we obtain the dissipation inequality (6.20)

$$S(x(t_2)) - S(x(t_1)) \le \int_{t_1}^{t_2} \frac{1}{2}(\gamma^2\|q(t)\|^2 - \|x(t)\|^2)\,dt$$

Now we take the time interval $[t_1, t_2]$ as $[0, T]$ and rearrange the above inequality, this becomes

$$S(x(T)) \le \gamma^2 \int_0^T \|q(t)\|^2\,dt + S(x(0)) - \int_0^T \|x(t)\|^2\,dt$$

By the fact $S(x) \ge 0$, and hence $S(x(T)) \ge 0$. We obtain the L_2-gain property (6.21):

$$\int_0^T \|x(t)\|^2\,dt \le \gamma^2 \int_0^T \|q(t)\|^2\,dt + S(x(0))$$

This completes the proof of Theorem 6.2.

We may reduce the construction of a solution to Hamilton-Jacobi equation (6.19) to solve a system of ordinary differential equations, as we did for the isolated signal intersection. This method is commonly known as the method of bicharacteristics in the nonlinear H_∞-control literature (see [van der Schaft 1996] and [Ball,Day and Kachroo 1997]).

6.7 Conclusion

Apart from the delay terms appearing in the case of (6.7) and (6.9), common features of all problems in this class of queueing networks which are not addressed in the standard nonlinear H_∞ control literature up to now are: (1) boundary dynamics, i.e., discontinuities in the dynamics caused by the constraint that the state vector remain in the physically prescribed constraint set (see [Dupuis, Ishii and Soner 1990] more on this feature) and (2) a finite admissible control set. It will be a challenge to develop some effective numerical algorithms for solving such complicated dynamic systems as (6.4), (6.9) and (6.10).

Chapter 7

Conclusions and Future

Traffic congestion has been cited as the major problem in today's traffic management. It has far-reaching economic, social and political effects. Intelligent Transportation Systems research and development programs have been assigned the task of developing sophisticated techniques and counter-measures to reduce traffic congestion to manageable levels. This goal makes it necessary for traffic signalization models to capture the dynamic nature of the traffic flow through intersections and networks. This thesis has made some progress toward the goal. We developed state space models for both isolated multi-phase intersections and a class of queueing networks, formulated H_∞ problems for the control systems with persistent disturbances, presented the projection dynamics aspects of the problem to account for the constraints on the state variables, formally to study this problem as a hybrid system; and also derived traffic-actuated feedback control laws for the multi-phase intersections.

A dynamical system has finite L_2-gain if it is dissipative in some sense. Therefore, the H_∞-control problem turns to designing a controller such that the resulting closed loop system is dissipative, and correspondingly there exists a storage function.

For the nonlinear traffic system, the theory of dissipative systems provides an interesting vehicle to explore the robust feedback solutions. The energy storage function, which measures the energy that may be extracted from the system, serves as a Lyapunov function to imply stability. This function satisfies Hamilton-Jacobi inequality. The existence of such a function further implies that the closed-loop system has L_2-gain less than or equal to a pre-specified constant level γ. Solving the nonlinear state space H_∞ problem refers to finding a control $u = l(x)$ which will guarantee a smallest bounded L_2-gain for the system.

Though we have mathematically presented a robust feedback solution for the traffic signalization, our controller needs some field tests before any physical implementation.

In conventional systems, full realization of benefits of traffic signalization depends on the frequent updating of signal plans to optimize traffic flow, but it is labor intensive and costly. Therefore, it is nice to have "adaptive" property in the signal controller, so that the system has the capability to automatically change the signal timing in response to both short term and longer term variations in traffic conditions. The adaptive system can save a lot of human and financial resources to update the system's databases. That is, an adaptive control strategy promises significant improvement in currently used traffic signal control mechanisms to provide:

- The best possible signal timing in all types of traffic situations, and
- Update signal timings with limited effort.

Of course, the big challenge is to design practical adaptive controllers which provide good attenuation of disturbances, are robust to unmodeled dynamics, and are computationally tractable.

Bibliography

[Aubin 1991] J. P. Aubin, *Viability Theory*, Birkhauser, Boston, 1991.

[Akcelik 1985] R. Akcelik, "SIDRA-2 for Traffic Signal Design." *Traffic Engineering and Control 26*, 1985, pp. 256-261.

[Akcelik 1985] R. Akcelik, "Estimation of Green Times and Cycle Time for Vehicle-Actuated Signals", *Transportation Research Record 1457*, 1994, pp. 63-72.

[Ball, Day, Kachroo and Yu 1997] J.A. Ball, M. V. Day, P. Kachroo and T. Yu, "Hybrid Binary H_∞-Control: An Isolated Traffic Intersection", preprint, Virginia Tech, 1997.

[Ball, Day and Kachroo 1997] J.A. Ball, M. V. Day, and P. Kachroo, "Hybrid H_∞-Control for a Multiphase Traffic Intersection", in preparation, Virginia Tech, 1997.

[Ball, Helton and Walker 1993] J.A. Ball, J.W. Helton, and M.L. Walker, "H_∞ Control for Nonlinear Systems with Output Feedback", *IEEE Trans. on Automatic Control*, Vol. 38, April 1993, pp. 546-559.

[Ball and Helton 1996] J.A. Ball and J.W. Helton, "Viscosity solutions of Hamilton-Jacobi equations arising in nonlinear H_∞-control", (Summary), *J. Math. Systems, Estimation, and Control* Vol. 6, 1996, pp. 109-112.

[Basar and Bernhard 1991] T. Basar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, Birkhauser, 1991.

[Cremer and Fleischmann 1987] M. Cremer and S. Fleischmann, "Traffic Responsive Control of Freeway networks by State Feedback Approach", In *N. H. Gartner and N. H. M. Wilson (EDS), Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, Elsevier, New York, 1987.

[Coddington and Levinson 1955] E. A. Coddington, N. L. Levinson, *Theory of ordinary differential equations*, New York, McGraw-Hill, 1955.

[Day 1997] M.V. Day, "On Lagrange manifolds and viscosity solutions", preprint, Virginia Tech, 1997.

[Drew 1963] Donald R. Drew, "Design and Signalization of High-Type Facilities", *Traffic Engineering*, Vol. 33, No. 7, 1963, pp. 17-25.

[Dunne and Potts 1964] M.C. Dunne and R.B. Potts, "Algorithm for Traffic Control", *Operation Research*, Vol. 12, 1964, pp. 870-881.

[Dupuis, Ishii and Soner 1990] P. Dupuis, H. Ishii and H.M. Soner, A viscosity solution approach to the asymptotic analysis of queueing systems, *The Annals of Probability* 18 (1990), 226-255.

[FHWA 1996] *Traffic Control Systems Handbook*, Publication Number FHWA-SA-95-032, February 1996.

[Filippov 1979] A.F.Filippov, "Differential Equations with Second Members Discontinuous on Intersecting Surfaces", *Differentsial'nye Uravneniya*, vol.15, no.10, pp.1814-1832,1979.

[Filippov 1964] A. F. Filippov, "Differential equations with discontinuous right hand sides", *Mathematicheskii Sbornek*, 51, 1 (1960), in Russian. Translated in English, *Am. Math. Soc. Trans.*, 62, 199, 1964.

[Gartner 1983] N.H. Gartner, "OPAC: A Demand-Responsive Strategy for Traffic Control." *Transportation Research Record 906*, TRB, Washington, D.C. 1983.

[Gazis 1964] D.C. Gazis, "Optimum Control of a System of Oversaturated Intersections", *Operation Research*, Vol. 12, 1964, pp. 815-831.

[Gazis 1974] D.C. Gazis, "Modeling and Optimal Control of Congested Transportation Systems," *Network 4*, 1974, pp. 113-124.

[Gutman 1979] S. Gutman, "Uncertain Dynamical Systems - A Lyapunov Min-Max Approach", *IEEE Trans. on Aut. Cont.*, vol.AC-24, no.3, pp. 437-443, 1979.

[HCM 1985] Highway Capacity Manual, "Special Report", *Transportation Research Board*, Washington, D.C. 1985.

[Heydecker 1987] B. Heydecker, "Uncertainty and Variability in Traffic Signal Calculations", *Transportation Research, -B* Vol. 21B, No. 1, 1987, pp. 79-85.

[Isidori 1995] A. Isidori and W. Kang, "H_∞ Control via Measurement Feedback for General Nonlinear Systems", *IEE Trans. on Automatic Control*, vol. 40, No. 3, March 1995.

[James 1993] M. James, "A partial differential inequality for dissipative nonlinear systems, *Systems & Control Letters* Vol. 21, 1993, pp. 315-320.

[Kessmann and Ganslaw 1973] Kessmann and Ganslaw, "Control Strategies of the Traffic Adaptive Network Signal Timing Program (TANSTP)", *Notes on paper presented at Annal FCP Research Progress Review Meeting,* Annapolis, Maryland, August, 1973.

[Lin 1982] F.B. Lin, "Estimation of Average Phase Duration for Full-Actuated Signals". *Transportation Research Record 881*, 1982, pp. 65-72.

[Lindley 1987] J. A. Lindley, "Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions", *ITE Journal*, January 1987.

[Longley 1968] D. Longley, "A Control Strategy for a Congested Computer-Controlled Traffic Network", *Transportation Research*, Vol. 2, 1968, pp. 391-408.

[McShane and Roess 1990] W. R. McShane and R. P. Roess, *Traffic Engineering*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1990.

[McHaney 1991] R. McHaney, *Computer Simulation: A Practical Perspective*, Academic Press, Inc., San Diego, CA, 1991.

[Michalopoulos and Stephanopoulos 1978] P. G. Michalopoulos and G. Stephanopoulos, "Optimal Control of Oversaturated Intersection." *Traffic Engineering and Control*, Vol. 19, No. 5, 1978.

[Miller 1965] A.J. Miller, "A Computer Control System for Traffic Networks", *Proceedings, Second International Symposium on the Theory of Road Traffic Flow*, Organization for Economic Co-operation and Development, pp. 200-220., Paris, 1965.

[Morin et al. 1991] J. M. Morin, P. Gower, M. Papageorgiou and A. Messmer, "Motorway Networks, Modeling and Control", DRIVE Project No. V 1035, *DRIVE conference: Advanced telematics in road transport*, 1991.

[Nairn 1984] R. J. Nairn, "TRANSTEP". *Proc. Seminar on Microcomputers for the Transport Industry,* Melbourne, Pak Poy and Kneebone Pty. Ltd. Adelaide, Australia, 1984.

[Nairn and Partners 1986] R. J. Nairn and Partners, "Intersection Simulation Model (INSECT)." *Department of Main Roads,* NSW, Australia, 1986.

[OECD 1981] *Traffic Control in Saturated Conditions*, Organization for Economic Cooperation and Development, Paris, 1981.

[Paden and Satry 1987] B. Paden and S. Sastry, "A Calculus for Computing Filippov's Differential Inclusion with Application to the Variable Structure Control of Robot Manipulators", *IEEE Trans. Circ. Syst.*, vol. CAS-34, no.1, pp. 73-82, 1987.

[Papageorgiou 1990] M. Papageorgiou, "Dynamic Modeling, Assignment, and Route Guidance in Traffic Networks", *Transportation Research, Part B* Vol. 24b, No. 6, 1990, pp. 471-495.

[Payne 1971] H. J. Payne, "Models of freeway traffic and control". *Simulation council*, 1979, pp. 51-61.

[Payne 1979] H. J. Payne, "FREFLO: a Macroscopic Simulation Model of Freeway Traffic". *Transportation Research Record 722*, 1979, pp. 68-77.

[Payne and Thompson 1974] H. J. Payne and W. A. Thompson, "Allocation of Freeway Ramp Meeting Volumes to Optimize Corridor Performance", *IEEE Trans. and Automatic Control AC-19*, 1974, pp. 177-186.

[Pignataro 1975] L.J. Pignataro, W.R. McShane, K. W. Crowley, B. Lee and T.W. Casey, "Traffic Control in Oversaturated Street Network", *National Cooperate Highway Research Program Report*, No. 194, Washington, D.C., 1975.

[Puri and Varaiya 1994] A. Puri and P. Varaiya, "Decidability of hybrid systems with retangular differential inclusions", *Proc. 6th Workshop Computer-Aided Verification*, 1994.

[Reeves 1984] C. M. Reeves, "Complexity Analyses of Event Set Algorithms", *The Computer Journal*, 27(1), 1984, pp. 72-79.

[Rogness and Messer 1983]] R. O. Rogness and C. J. Messer, " A Heuristic Programming Approach to Arterial Signal Timing". *Transportation Research Record 906*, 1983, pp. 67-74.

[Semmens 1985] M. C. Semmens, "PICADY 2: An Enhanced Program to Model Capacities, queues and Delays at Major/Minor Priority Junctions", *Transport and Road Research Laboratory*, Research Report 36, UK, 1985.

[Sheffi 1985] Y. Sheffi, "Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods", *PRENTICE-HALL, INC.*, Englewood Cliffs, New Jersey, 1985.

[Staunton 1976] M.M. Staunton, "Vehicle Actuated Signal Controls for Isolated Lacation". *An Foras Forbartha*, Dublin, Ireland, 1976.

[UMTA 1985] "Microcomputers in Transportation: Software Source Book." *Urban Mass Transit Association*, 1985, US Department of Transportation, Washington, DC.

[van der Schaft 1996] A. J. van der Schaft, *L_2-Gain and Passivity Techniques in Nonlinear Control*, Lecture Notes in Control and Information Sciences #218, Springer, 1996.

[Van Schuppen, 1992] J.H. Van Schuppen, "Routine of freeway traffic - a discrete-time state space model and routing problem", *Report BS-R9232*, CWI, Amsterdam, 1992.

[Shevitz and Paden 1994] D. Shevitz, and B. Paden, "Lyapunov Stability Theory of Nonsmooth Systems", *IEEE Trans. on Aut. Cont.*, vol.39, no.9, pp. 1910-1914, 1994.

[Shiva 1988] S. G. Shiva, *Introduction to Logic Design*, Scott, Foresman and Company, 1988.

[Utkin 1978] V. I. Utkin, *Sliding Modes and Their Application to Variable Structure Systems*, MIR Publishers, Moscow, 1978.

[Webster 1958] F.V. Webster, "Traffic Signal Settings", *Road Research Technical Paper*, No. 39, London, 1958.

[Willems 1972] J.C. Willems, "Dissipative dynamic systems - Part I: General Theory", *Arch. Rational Mechanics and Analysis*, vol. 45, 1972, pp. 321-351.

[Young et al. 1989] W. Young, M. A. P. Taylor and P. G. Gipps, "Microcomputers in Traffic Engineering", *Research Studies Press LTD.* Taunton, Somerset, England, 1989.

[Zames 1981] G. Zames, "Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses", *IEEE Transactions on Automatic Control*, Vol. 26, 1981, pp. 301-320.

VITA

Tungsheng Yu entered this world on November 11, 1964 in Ningbo, China. He survived from a truck accident when he was six. Two years later he went to school. He also worked as a part-time farmer with his brother to support his family for six years under the rule of Chairman Mao. In September 1982 he went to college and graduated with a MS degree in Applied Mathematics from Peking University in July 1989. After graduation he came back to his hometown and worked as system engineer at Ningbo Computer Center, Ningbo, China. In July 1990 Mr. Yu arrived at USA and joined a PhD program in Mathematics at Virginia Polytechnic Institute and State University (i.e. Virginia Tech). In May 1991 he registered as a graduate in Computer Science at Virginia Tech, where he received a MS degree in Computer Science in September 1992. Then he worked as software engineer for five years at VTLS Inc. In September 1997 he moved into Ashburn Village — a wonderful new land and works as computer system consultant at AT&T. After many years of hard work, he received a PhD in Mathematics from Virginia Tech in 1997.