A. A. Agrachev A. S. Morse E. D. Sontag
H. J. Sussmann V. I. Utkin

NONLINEAR AND OPTIMAL CONTROL THEORY

Lectures given at
the C.I.M.E. Summer School
held in Cetraro, Italy,
June 19-29, 2004

Editors: P. Nistri and G. Stefani

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Contents

Lecture notes on Logically Switched Dynamical Systems

A. Stephen Morse *

Yale University, USA
morse@sysc.eng.yale.edu

Introduction

The subject of logically switched dynamical systems is a large one which overlaps with may areas including hybrid system theory, adaptive control, optimal control, cooperative control, etc. Ten years ago we presented a lecture, documented in [1], which addressed several of the areas of logically switched dynamical systems which were being studied at the time. Since then there have been many advances in many directions, far to many too adequately address in these notes. One of the most up to date and best written books on the subject is the monograph by Liberzon [2] to which we refer the reader for a broad but incisive perspective as well an extensive list of references.

In these notes we will deal with two largely disconnected topics, namely switched adaptive control {sometimes called supervisory control} and "flocking" which is about the dynamics of reaching a consensus in a rapidly changing environment. In the area of adaptive control we focus mainly on one problem which we study in depth. Our aim is to give a thorough analysis under realistic assumptions of the adaptive version of what is perhaps the most important design objective in all of feedback control, namely set-point control of a single-input, single output process admitting a linear model. While the non-adaptive version the set-point control problem is very well understood and has been so for more than a half century, the adaptive version still is not because there is no credible counterpart in an adaptive context of the performance theories which address the non-adaptive version of the problem. In fact, even just the stabilization question for the adaptive version of the problem did not really get ironed out until ten years ago, except under unrealistic assumptions which ignored the effects of noise and/or un-modelled dynamics.

As a first step we briefly discuss the problem of adaptive disturbance rejection. Although the switching logic we consider contains no logic or discrete

* This research was supported by the US Army Research Office, the US National Science Foundation and by a gift from the Xerox Corporation

event sub-system, the problem nonetheless sets the stage for what follows. One of the things which turns out {in retrospect} to have impeded progress with adaptive control has been the seemingly benign assumption that the parameters of the {nominal} model of the process to be controlled are from a *continuum* of possible values. In Chapter 4 we briefly discuss the unpleasant consequences of this assumption and outline some preliminary ideas which might be used to deal with them.

In Chapter 5 we turn to detailed discussion of a switched adaptive controller capable of causing the output of an imprecisely modelled process to approach and track a constant reference signal. The material in this chapter provides a clear example of what is meant by a logically switched dynamical system. Finally in Chapter 6 we consider several switched dynamical systems which model the behavior of a group of mobile autonomous agents moving in a rapidly changing environment using distributed controls. We begin with what most would agree is the quintessential problem in the area of switched dynamical systems.

1 The Quintessential Switched Dynamical System Problem

The quintessential problem in the area of switched dynamical systems is this: Given a compact subset $\mathcal{P}$ of a finite dimensional space, a parameterized family of $n \times n$ matrices $\mathcal{A} = \{A_p : p \in \mathcal{P}\}$, and a family $\mathcal{S}$ of piecewise-constant switching signals $\sigma : [0, \infty) \to \mathcal{P}$, determine necessary and sufficient conditions for A_σ to be exponentially stable for every $\sigma \in \mathcal{S}$. There is a large literature on this subject. Probably its most comprehensive treatment to date is in the monograph [2] by Liberzon mentioned before. The most general version of the problem is known to be undecidable [3]. These notes deal with two special versions of this problem. Each arises in a specific context and thus is much more structured than the general problem just formulated.

1.1 Dwell-Time Switching

In the first version of the problem, $\mathcal{S}$ consists of all switching signals whose switching times are separated by τ_D times unit where τ_D is a pre-specified positive number called a *dwell time*. More precisely, $\sigma \in \mathcal{S}$ is said to have dwell time τ_D if and only if σ switches values at most once, or if it switches more that once, the set of time differences between any two successive switches is bounded below τ_D. In Section 5.1 we will encounter a switching logic which generates such signals.

Note that the class $\mathcal{S}$ just defined contains constant switching signal $\sigma(t) = p,\ t \geq 0$ for any value of $p \in \mathcal{P}$. A necessary condition for A_σ to be exponentially stable for every $\sigma \in \mathcal{S}$, is therefore that each $A_p \in \mathcal{A}$ is exponentially stable. In other words, if A_σ to be exponentially stable for every

$\sigma \in \mathcal{S}$, then for each $p \in \mathcal{P}$ there must exist non-negative numbers t_p and λ_p, with λ_p positive such that $|e^{A_p t}| \leq e^{\lambda_p(t_p - t)}, \quad t \geq 0$. Here and elsewhere through the end of Chapter 5, the symbol $|\cdot|$ denotes any norm on a finite dimensional linear space. It is quite easy to show by example that this condition is not sufficient unless τ_D is large. An estimate of how large τ_D has to be in order to guarantee exponential stability, is provided by the following lemma.

Lemma 1.1. *Let $\{A_p : p \in \mathcal{P}\}$ be a set of real, $n \times n$ matrices for which there are non-negative numbers t_p and λ_p with λ_p positive such that*

$$|e^{A_p t}| \leq e^{\lambda_p(t_p - t)}, \quad t \geq 0 \tag{1.1}$$

Suppose that τ_D is a finite number satisfying

$$\tau_D > t_p, \quad p \in \mathcal{P} \tag{1.2}$$

For any switching signal $\sigma : [0, \infty) \to \mathcal{P}$ with dwell time τ_D, the state transition matrix of A_σ satisfies

$$|\Phi(t, \mu)| \leq e^{\lambda(T - (t-\mu))}, \ \ \forall\, t \geq \mu \geq 0 \tag{1.3}$$

where λ is a positive number defined by

$$\lambda = \inf_{p \in \mathcal{P}} \left\{ \lambda_p \left(1 - \frac{t_p}{\tau_D} \right) \right\} \tag{1.4}$$

and

$$T = \frac{2}{\lambda} \sup_{p \in \mathcal{P}} \{\lambda_p t_p\} \tag{1.5}$$

Moreover,

$$\lambda \in (0, \lambda_p], \quad p \in \mathcal{P} \tag{1.6}$$

The estimate given by this lemma can be used to make sure that the "slow switching assumption" discussed in Section 5.2 is satisfied.

Proof of Lemma 1.1: Since $\mathcal{P}$ is a closed, bounded set, $\sup_{p \in \mathcal{P}} t_p < \infty$. Thus a finite τ_D satisfying (1.2) exists. Clearly $\lambda_p(1 - \frac{t_p}{\tau_D}) > 0,\ p \in \mathcal{P}$. From this and the definition of λ it follows that (1.6) holds and that

$$e^{\lambda_p(t_p - \tau_D)} \leq e^{-\lambda \tau_D}, \quad p \in \mathcal{P}$$

This and (1.1) imply that for $t \geq \tau_D$

$$\begin{aligned} |e^{A_p t}| &\leq e^{\lambda_p(t_p - t)} \\ &= e^{\lambda_p(t_p - \tau_D)} e^{-\lambda_p(t - \tau_D)} \\ &\leq e^{-\lambda \tau_D} e^{-\lambda_p(t - \tau_D)} \\ &\leq e^{-\lambda \tau_D} e^{-\lambda(t - \tau_D)} \\ &\leq e^{-\lambda t}, \quad t \geq \tau_D,\ p \in \mathcal{P} \end{aligned} \tag{1.7}$$

It also follows from (1.1) and the definition of T that

$$|e^{A_p t}| \leq e^{\lambda(\frac{T}{2}-t)}, \quad t \in [0, \tau_D), \quad p \in \mathcal{P} \tag{1.8}$$

Set $t_0 = 0$ and let $t_1, t_2 \ldots$ denote the times at which σ switches. Write p_i for the value of σ on $[t_{i-1}, t_i)$. Note that for $t_{j-1} \leq \mu \leq t_j \leq t_i \leq t \leq t_{i+1}$,

$$\Phi(t,\mu) = e^{A_{p_{i+1}}(t-t_i)} \left(\prod_{q=j+1}^{i} e^{A_{p_q}(t_q - t_{q-1})} \right) e^{A_{p_j}(t_j - \mu)}$$

In view of (1.7) and (1.8)

$$\begin{aligned} |\Phi(t,\mu)| &\leq |e^{A_{p_{i+1}}(t-t_i)}| \left(\prod_{q=j+1}^{i} |e^{A_{p_q}(t_q - t_{q-1})}| \right) |e^{A_{p_j}(t_j - \mu)}| \\ &\leq e^{\lambda(\frac{T}{2}-(t-t_i))} \left(\prod_{q=j+1}^{i} e^{-\lambda(t_q - t_{q-1})} \right) e^{\lambda(\frac{T}{2}-(t_j-\mu))} \\ &= e^{\lambda(T-(t-\mu))} \end{aligned}$$

On the other hand, for $i > 0$, $t_{i-1} \leq \mu \leq t \leq t_i$, (1.8) implies that

$$|\Phi(t,\mu)| \leq e^{\lambda(\frac{T}{2}-(t-\mu))} \leq e^{\lambda(T-(t-\mu))}$$

and so (1.3) is true.■

Input-Output Gains of Switched Linear Systems

In deriving stability margins and systems gains for the supervisory control systems discussed in Chapter 5 we will make use of induced "gains" (i.e. norms) of certain types of switched linear systems. Quantification of stability margins and the devising of a much needed performance theory of adaptive control, thus relies heavily on our ability to characterize these induced gains. In this section we make precise what types of induced gains we are referring to and we direct the reader to some recent work aimed at their characterization.

To begin, suppose that $\{(A_p, B_p, C_p, D_p) : p \in \mathcal{P}\}$ is a family of coefficient matrices of m-input, r-output, n-dimensional, exponentially stable linear systems. Then any $\sigma \in \mathcal{S}$ determines a switched linear system of the form

$$\Sigma_\sigma \stackrel{\Delta}{=} \left\{ \begin{array}{l} \dot{x} = A_\sigma x + B_\sigma u \\ y = C_\sigma x + D_\sigma u \end{array} \right\} \tag{1.9}$$

Thus if $x(0) \stackrel{\Delta}{=} 0$, then $y = Y_\sigma \circ u$, where Y_σ is the input-output mapping

$$u \longmapsto \int_0^t C_{\sigma(t)} \Phi(t,\tau) D_{\sigma(\tau)} B_\sigma(\tau) u(\tau) d\tau + D_\sigma u,$$

and Φ is the state transition matrix of A_σ. Let prime denotes transpose and, for any integrable, vector-valued signal v on $[0,\infty)$, let $||\cdot||$ denotes the two-norm

$$||v|| \stackrel{\Delta}{=} \sqrt{\int_0^\infty v'(t)v(t)dt}$$

The *input-output gain* of Σ_σ is then the induced two-norm

$$\gamma(\sigma) \stackrel{\Delta}{=} \inf\{g : ||Y \circ u|| \leq g||u||,\ \forall u \in \mathcal{L}_2\}$$

where $\mathcal{L}_2$ is the space of all signals with finite two-norms. Define the gain $\mathfrak{g}$ of the *multi-system* $\{(A_p, B_p, C_p, D_p), p \in \mathcal{P}\}$ to be

$$\mathfrak{g} \stackrel{\Delta}{=} \sup_{\sigma \in \mathcal{S}} \gamma(\sigma)$$

Thus $\mathfrak{g}$ is the worst case input-output gain of (1.9) as σ ranges over all switching signals in $\mathcal{S}$. Two problems arise:

1. Derive conditions in terms of τ_D and the multi-system $\{(A_p, B_p, C_p, D_p), p \in \mathcal{P}\}$ under which $\mathfrak{g}$ is a finite number.
2. Assuming these conditions hold, characterize $\mathfrak{g}$ in terms of $\{(A_p, B_p, C_p, D_p), p \in \mathcal{P}\}$ and τ_D.

The first of the two problems just posed implicitly contains as a subproblem the quintessential switched dynamical system problem posed at the beginning of this section. A sufficient condition for $\mathfrak{g}$ to be finite is that τ_D satisfies condition (1.2) of Lemma 1.1. For the second problem, what would be especially useful would be a characterization of $\mathfrak{g}$ which is coordinate-independent; that is a characterization which depends only on the transfer matrices $C_p(sI - A_p)^{-1}B_p + D_p,\ p \in \mathcal{P}$ and not on the specific realizations of these transfer matrices which define $\{(A_p, B_p, C_p, D_p), p \in \mathcal{P}\}$. For example, it is reasonable to expect that there might be a characterization of $\mathfrak{g}$ in terms of the $\mathcal{H}^\infty$ norms of the $C_p(sI - A_p)^{-1}B_p + D_p,\ p \in \mathcal{P}$, at least for τ_D sufficiently large.

The problem of characterizing $\mathfrak{g}$ turns out to be a good deal more difficult that one might at first suspect, even if all one wants is a characterization of the limiting value of $\mathfrak{g}$ as $\tau_D \to \infty$ [4]. In fact, contrary to intuition, one can show by example that this limiting value may, in some cases, be strictly greater than the supremum over $\mathcal{P}$ of the $\mathcal{H}_\infty$ norms of the $C_p(sI - A_p)^{-1}B_p + D_p$. We refer the interested reader to [4] for a more detailed discussion of this subject. It is results along these lines which will eventually lead to a bona fide performance theory for adaptive control.

1.2 Switching Between Stabilizing Controllers

In many applications, including those discussed in Chapter 5, the matrix A_σ arises within a linear system which models the closed loop connection consisting of fixed linear system in feedback with a switched linear system. For

example, it is possible to associate with a given family of controller transfer functions $\{\kappa_p : p \in \mathcal{P}\}$ together with a given process model transfer function τ, a switched control system of the form shown in Figure 1 where $\mathbb{C}_\sigma$ is a switched controller with instantaneous transfer function κ_σ.

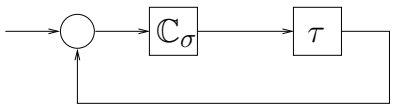

Fig. 1. A Switched Control System

Not always appreciated, but nonetheless true is the fact that the input-output properties of such a system depends on the specific realizations of the κ_p. Said differently, it is really not possible to talk about switched linear systems from a strictly input-output point of view. A good example, which makes this point, occurs when for each $p \in \mathcal{P}$, the closed loop systems shown in Figure 2 is stable.

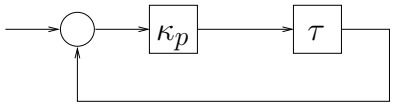

Fig. 2. A Linear Control System

Under these conditions, the system shown in Figure 1 turns out to be exponentially stable for every possible piecewise continuous switching signal, no matter how fast the switching, but only for certain realizations of the κ_p [5]. The key idea is to first use a Youla parameterization to represent the entire class of controller transfer functions and second to realize the family in such a way so that what's actually being switched within $\mathbb{C}_\sigma$ are suitably defined realizations of the Youla parameters, one for each κ_p. We refer the reader to [5] for a detailed discussion of this idea.

1.3 Switching Between Graphs

The system just discussed is a switched dynamical system because the controller within the system is a switched controller. Switched dynamical systems can arise for other reasons. An interesting example of this when the overall system under consideration models the motions of a group of mobile autonomous agents whose specific movements are governed by strategies which depend on

the movements of their nearby agents. A switched dynamical model can arise in this context because each agent's neighbors may change over time. What's especially interesting about this type of system is the interplay between the underlying graphs which characterize neighbor relationships, and the evolution of the system over time. Chapter 6 discusses in depth several example of this type of system.

2 Switching Controls with Memoryless Logics

2.1 Introduction

Classical relay control can be thought of as a form of switching control in which the logic generating the switching signal is a memoryless system. A good example of this is the adaptive disturbance rejector devised by I. M. Lie Ying at the High Altitude Control Laboratory at the Tibet Institute of Technology in Lhasa. Ying's work provides a clear illustration of what the use of switching can accomplish in a control setting, even if there is no logic or memory involved.

2.2 The Problem

In [6], Ying gives a definitive solution to the long-standing problem of constructing an adaptive feedback control for a one-dimensional siso linear system with an unmeasurable, bounded, exogenous disturbance input so as to cause the system's output to go to zero asymptotically. More specifically he considered the problem of constructing an adaptive feedback control $u = f(y)$ for the one-dimensional linear system

$$\dot{y} = -y + u + d$$

so as to cause the system's output y to go to zero no matter what exogenous disturbance $d : [0, \infty) \to \mathbb{R}$ might be, so long as d is bounded and piecewise-continuous. Up until the time of Ying's work, solutions to this long standing problem had been shown to exist only under the unrealistic assumption that d could be measured [7]. Ying made no such assumption.

2.3 The Solution

The adaptive control he devised is described by the equations

$$\begin{aligned} u &= -k\sigma(y) \\ \dot{k} &= |y| \end{aligned}$$

where

$$\sigma(y) = \begin{cases} 1 & y \geq 0 \\ -1 & y < 0 \end{cases} \tag{2.1}$$

The closed-loop system is thus

$$\dot{y} = -y - k\sigma(y) + d \tag{2.2}$$
$$\dot{k} = |y| \tag{2.3}$$

2.4 Analysis

Concerned that skeptical readers might doubt the practicality of his idea, Ying carried out a full analysis of the system. Here is his reasoning.

To study this system's behavior, let b be any positive number for which

$$|d(t)| \leq b, \quad t \geq 0 \tag{2.4}$$

and let V denote the Lyapunov function

$$V = \frac{y^2}{2} + \frac{(k-b)^2}{2} \tag{2.5}$$

In view of the definition of σ in (2.1), the rate of change of V along a solution to (2.2) and (2.3) can be written as

$$\begin{aligned} \dot{V} &= \dot{y}y + \dot{k}(k-b) \\ &= -y^2 - k|y| + dy + |y|(k-b) \end{aligned}$$

Since (2.4) implies that $dy \leq b|y|$, $\dot{V}$ must satisfy

$$\dot{V} \leq -y^2 - k|y| + b|y| + |y|(k-b)$$

Therefore

$$\dot{V} \leq -y^2$$

From this and (2.5) it follows that y and k are bounded and that y has a bounded $\mathcal{L}^2$ norm. In addition, since (2.2) implies that $\dot{y}$ is bounded, it must be true that $y \to 0$ as $t \to \infty$ {cf. [8]} which is what is desired. The practical significance of this result, has been firmly established by computer simulation performed be numerous graduate students all over the world.

3 Collaborations

Much of the material covered in these notes has appeared in one form or another in published literature. There are however several notable exceptions in Chapter 6. Among these are the idea of composing directed graphs and the interrelationships between rooted graphs, Sarymsakov graphs, and neighbor shared graphs discussed in Section 6.1. The entire section on measurement

delays {Section 6.3} is also new. All of the new topics addressed in Chapter 6 were developed in collaboration with Ming Cao and Brian Anderson. Daniel Spielman also collaborated with us on most of the convergence results and Jia Fang helped with the development as well. Most of this material will be published elsewhere as one or more original research papers.

4 The Curse of the Continuum

Due to its roots in nonlinear estimation theory, parameter identification theory generally focuses on problems in which unknown model parameters are assumed to lie within given continuums. Parameter-adaptive control, which is largely an outgrowth of identification theory, also usually addresses problems in which unknown process model parameters are assumed to lie within continuums. The continuum assumption comes with a large price tag because a typical parameter estimation problem over a continuum is generally not tractable unless the continuum is convex and the dependence on parameters is linear. These practical limitations have deep implications. For example, a linearly parameterized transfer matrix on a convex set can easily contain points in its parameter spaces at which the transfer matrix has an unstable pole and zero in common. For nonlinear systems, the problem can be even worse - for those process model parameterizations in which parameters cannot be separated from signals, it is usually impossible to construct a finite-dimensional multi-estimator needed to carry out the parameter estimatimation process. We refer to these and other unfortunate consequences of the continuum assumption as the *Curse of the Continuum.* An obvious way to avoid the curse, is to formulate problems in such a way so that the parameter space of interest is finite or perhaps countable. But many problems begin with parameter spaces which are continuums. How is one to reformulate such a problem using a finite parameter space, without serious degradation in expected performance? How should a parameter search be carried out in a finite parameter space so that one ends up with a provably correct overall adaptive algorithm? It is these questions to which this brief chapter and Chapter 5 are addressed.

4.1 Process Model Class

Let $\mathbb{P}$ be a process to be controlled and suppose for simplicity that $\mathbb{P}$ is a siso system admitting a linear model. Conventional linear feedback theory typically assumes that $\mathbb{P}$'s transfer function lies in a known open ball

$$\mathbb{B}(\nu, r)$$

of radius r centered at nominal transfer function ν in a metric space $\mathcal{T}$. In contrast, main-stream adaptive control typically assumes $\mathbb{P}$'s transfer function lies in a known set of the form

$$\mathcal{M} = \bigcup_{p \in \mathcal{P}} \mathbb{B}(\nu_p, r_p)$$

where $\mathcal{P}$ is a compact continuum within a finite dimensional space and $p \longmapsto r_p$ is at least bounded.

In a conventional non-adaptive control problem, a controller transfer function κ is chosen to endow the feedback system shown in Figure 3a with stability and other prescribed properties for each candidate transfer function $\tau \in \mathbb{B}(\nu, r)$. Control of $\mathbb{P}$ is then carried out by applying to $\mathbb{P}$ a controller $\mathbb{C}$ with transfer function κ as shown in Figure 3b.

(a) (b)

Fig. 3. Non–Adaptive Control

In the adaptive control case, for each $p \in \mathcal{P}$, controller transfer function κ_p is chosen to endow the feedback system shown in Figure 4a with stability and other prescribed properties for each $\tau \in \mathbb{B}(\nu_p, r_p)$. Adaptive control of $\mathbb{P}$ is then carried, in accordance with the idea of "certainty equivalence by using a parameter-varying controller or *multi-controller* $\mathbb{C}_\sigma$ with instantaneous transfer function κ_σ where σ is the index in $\mathcal{P}$ of the "best" current estimate of the ball within which $\mathbb{P}$'s transfer function resides[2].

Since $\mathcal{P}$ is a continuum and κ_p is to be defined for each $p \in \mathcal{P}$, the actual construction of κ_p is at best a challenging problem, especially if the construction is based on LQG or $\mathcal{H}^\infty$ techniques. Moreover, because of the continuum, the associated estimation of the index of the ball within which $\mathbb{P}$'s transfer function resides will be intractable unless demanding conditions are satisfied. Roughly speaking, $\mathcal{P}$ must be convex and the dependence of candidate process models on p must be linear. And in the more general case of nonlinear

[2] *Certainty equivalence* is a heuristic idea which advocates that the feedback controller applied to an imprecisely modelled process should, at each instant of time, be designed of the basis of a current estimate of what the process is, with the understanding that each such estimate is to be viewed as correct even though it may not be. The term is apparently due to Herbert Simon [9] who used in a 1956 paper [10] to mean something somewhat different then what's meant here and throughout the field of parameter adaptive control. The consequence of using this idea in an adaptive context is to cause the interconnection of the controlled process, the multi-controller and the parameter estimator to be detectable through the error between the output of the process and its estimate for every frozen parameter estimate – and this is true whether the three subsystems involved are linear or not [11].

(a) (b)

Fig. 4. Adaptive Control

process models, a certain separability condition [11, 12] would have to hold which models as simple as

$$\dot{y} = \sin(py) + u$$

fail to satisfy. The totality of these implications is rightly called *the curse of the continuum*.

Example

The following example illustrates one of the difficulties which arises as a consequence of the continuum assumption. Ignoring un-modelled dynamics, suppose that $\mathcal{M}$ is simply the parameterized family of candidate process model transfer functions

$$\mathcal{M} = \left\{ \frac{s - \frac{1}{6}(p+2)}{s^2 + ps - \frac{2}{9}p(p+2)} : \quad p \in \mathcal{P} \right\}$$

where $\mathcal{P} = \{p : -1 \leq p \leq 1\}$. Note that there is no transfer function in $\mathcal{M}$ with a common pole and zero because the polynomial function

$$s^2 + ps - \frac{2}{9}p(p+2)|_{s=\frac{1}{6}(p+2)}$$

is nonzero for all $p \in \mathcal{P}$. The parameterized transfer function under consideration can be written as

$$\frac{s - \frac{1}{6}(p+2)}{s^2 + ps + q}$$

where

$$q = -\frac{2}{9}p(p+2) \tag{4.1}$$

Thus $\mathcal{M}$ is also the set of transfer functions

$$\mathcal{M} = \left\{ \frac{s - \frac{1}{6}(p+2)}{s^2 + ps + q} : \quad (p,q) \in \mathcal{Q} \right\} \tag{4.2}$$

where $\mathcal{Q}$ is the two-parameter space

$$\mathcal{Q} = \{(p,q) : q + \frac{2}{9}p(p+2) = 0,\ -1 \leq p \leq 1\}$$

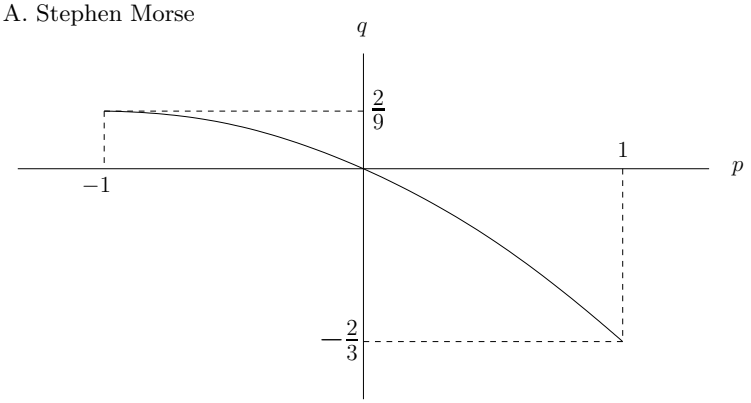

Fig. 5. Parameter Space $\mathcal{Q}$

The set of points in $\mathcal{Q}$ form a parabolic curve segment as shown in Figure 5.

Although the parameterized transfer function defining $\mathcal{M}$ in (4.2) depends linearly on p and q, the parameter space $\mathcal{Q}$ is not convex. Thus devising a provably parameter estimation algorithm for this parameterization would be difficult. In a more elaborate example of this type, where more parameters would be involved, the parameter estimation problem would typically be intractable.

There is a natural way to get around this problem, if the goal is identification as opposed to adaptive control. The idea is to embed $\mathcal{Q}$ in a larger parameter space which is convex. The smallest convex space $\bar{\mathcal{Q}}$ containing $\mathcal{Q}$ is the convex hull of $\mathcal{Q}$ as shown in Figure 6.

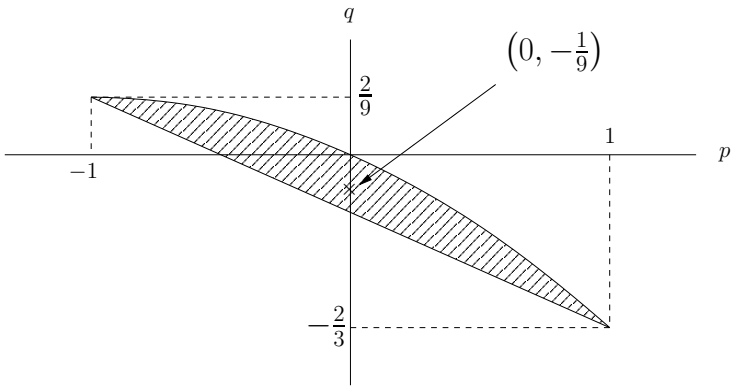

Fig. 6. Parameter Space $\bar{\mathcal{Q}}$

The corresponding set of transfer functions is

$$\bar{\mathcal{M}} = \left\{ \frac{s - \frac{1}{6}(p+2)}{s^2 + ps + q}, \quad (p,q) \in \bar{\mathcal{Q}} \right\}$$

It is easy to see that *any* linearly parameterized family of transfer functions containing $\mathcal{M}$ which is defined on a convex parameter space, must also contain $\bar{\mathcal{M}}$. While this procedure certainly makes tractably the problem of estimating parameters, the procedure introduces a new problem. Note that if the newly parameterized transfer function is evaluated at the point $(0, -\frac{1}{9}) \in \bar{\mathcal{Q}}$, what results is the transfer function

$$\left. \frac{s - \frac{1}{6}(p+2)}{s^2 + ps + q} \right|_{(p,\, q)=(0,-\frac{1}{9})} = \frac{s - \frac{1}{3}}{s^2 - \frac{1}{9}}$$

This transfer function has a common pole and zero in the right half plane at $s = \frac{1}{3}$. In summary, the only way to embed $\mathcal{M}$ in a larger set which is linearly parameterized on a convex parameter space, is to introduce candidate process model transfer functions which have right half plane pole-zero cancellations. For any such candidate process model transfer function τ, it is impossible to construct a controller which stabilizes the feedback loop shown in Figure 4a. Thus the certainty equivalence based approach we've outlined for defining $\mathbb{C}_\sigma$ cannot be followed here.

There is a way to deal with this problem if one is willing to use a different paradigm to construct $\mathbb{C}_\sigma$ [13]; the method relies on an alternative to certainty equivalence to define $\mathbb{C}_\sigma$ for values of σ which are "close" to points on parameter space at which such pole zero cancellations occur. Although the method is systematic and provably correct, the multi-controller which results is more complicated than the simple certainty-equivalence based multi-controller described above. There is another way to deal with this problem [14] which we discuss next.

4.2 Controller Covering Problem

As before let $\mathbb{P}$ be a siso process to be controlled and suppose that $\mathbb{P}$ has a model in

$$\mathcal{M} = \bigcup_{p \in \mathcal{P}} \mathbb{B}(\nu_p, r_p)$$

where $\mathbb{B}(\nu_p, r_p)$ is a ball of radius r_p centered at nominal transfer function ν_p in a metric space $\mathcal{T}$ with metric μ. Suppose in addition that $\mathcal{P}$ is a compact continuum within a finite dimensional space and $p \longmapsto r_p$ is at least bounded. Instead of trying re-parameterize as we did in the above example, suppose instead we try to embed $\mathcal{M}$ is a larger class of transfer functions which is the union of a *finite set* of balls in $\mathcal{T}$, each ball being small enough so that it can be adequately controlled with a single conventional linear controller. We pursue this idea as follows.

Let us agree to say that a finite set of controller transfer functions $\mathcal{K}$ is a *control cover* of $\mathcal{M}$ if for each transfer function $\tau \in \mathcal{M}$ there is at least one

transfer function $\kappa \in \mathcal{K}$ which endows the closed-loop system shown in Figure 7 with at least stability and possible other prescribed properties.

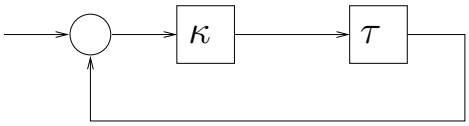

Fig. 7. Feedback Loop

The *controller covering problem* for a given $\mathcal{M}$ is to find a control cover, if one exists.

4.3 A Natural Approach

There is a natural way to try to solve the controller covering problem if each of the balls $\mathbb{B}(\nu_i, r_i)$ is small enough so that each transfer function in any given ball can be adequately controlled by the same fixed linear controller. In particular, if we could cover $\mathcal{M}$ with a finite set of balls - say $\{\mathbb{B}(\nu_p, r_p) : p \in \mathcal{Q}\}$ where $\mathcal{Q}$ is a finite subset of $\mathcal{P}$ - then we could find controller transfer functions $\kappa_p,\ p \in \mathcal{Q}$, such that for each $p \in \mathcal{Q}$ and each $\tau \in \mathbb{B}(\nu_p, r_p)$, κ_p provides the system shown in Figure 4a with at least stability and possibly other prescribed properties. The problem with this approach is that $\mathcal{M}$ is typically not compact in which case no such finite cover will exist. It is nevertheless possible to construct a finite cover of $\mathcal{M}$ using enlarged versions of some of the balls in the set

$$\{\mathbb{B}(\nu_p, r_p) : p \in \mathcal{P}\}$$

In fact, $\mathcal{M}$ can be covered by any one such ball. For example, for any fixed $q \in \mathcal{P}$, $\mathcal{M} \subset \mathbb{B}(\nu_p, s_q)$ where

$$s_q = \sup_{p \in \mathcal{P}} r_p + \sup_{p \in \mathcal{P}} \mu(\nu_p, \nu_q)$$

This can be checked by simply applying the triangle inequality. The real problem then is to construct a cover using balls which are small enough so that all transfer functions in any one ball can be adequately controlled by the same linear controller. The following lemma [15] takes a step in this direction.

Lemma 4.1. *If $\mathcal{Q}$ is a finite subset of $\mathcal{P}$ such that*

$$\{\nu_p : p \in \mathcal{P}\} \subset \bigcup_{q \in \mathcal{Q}} \mathbb{B}(\nu_q, r_q)$$

then

$$\mathcal{M} \subset \bigcup_{q \in \mathcal{Q}} \mathbb{B}(\nu_q, r_q + s_q)$$

where for $q \in \mathcal{Q}$,

$$s_q = \sup_{p \in \mathcal{P}_q} r_p$$

and $\mathcal{P}_q$ *is the set of all* $p \in \mathcal{P}$ *such that* $\nu_p \in \mathbb{B}(\nu_q, r_q)$

In other words, if we can cover the set of nominal process model transfer functions $\mathcal{N} = \{\nu_p : p \in \mathcal{P}\}$ with a finite set of balls from the set $\{\mathbb{B}(\nu_p, r_p) : p \in \mathcal{P}\}$, then by enlarging these balls as the lemma suggests, we can over $\mathcal{M}$. Of course for such a cover of $\mathcal{N}$ to exist, $\mathcal{N}$ must be compact. It is reasonable to assume that this is so and we henceforth do. But we are not yet out of the woods because the some of enlarged balls we end up with may still turn out to be too large to be robust stabilizable with linear controllers, even if each of the balls in the original set $\{\mathbb{B}(\nu_p, r_p) : p \in \mathcal{P}\}$ is. There is a different way to proceed which avoids this problem if certain continuity conditions apply. We discuss this next.

4.4 A Different Approach

Let us assume that

1. $p \longmapsto r_p$ is continuous on $\mathcal{P}$.
2. For each $p \in \mathcal{P}$ and each positive number ϵ_p there is a number δ_p for which
$$\mu(\nu_q, \nu_p) < \epsilon_p$$
whenever $|q - p| < \delta_p$ where $|\cdot|$ is the norm in the finite-dimensional space within which $\mathcal{P}$ resides.

The following lemma is from [14].

Lemma 4.2. *Let the preceding assumptions hold. For each function* $p \longmapsto s_p$ *which is positive on* $\mathcal{P}$*, there is a finite subset* $\mathcal{Q} \subset \mathcal{P}$ *such that*

$$\mathcal{M} \subset \bigcap_{q \in \mathcal{Q}} \mathbb{B}(\nu_q, r_q + s_q)$$

The key point here is that if the continuity assumptions hold, then it is possible to cover $\mathcal{M}$ with a finite set of balls which are arbitrarily close in size to the originals in the set $\{\mathbb{B}(\nu_p, r_p) : p \in \mathcal{Q}\}$. Thus if each of the original balls is robustly stabilizable with a linear controller, then so should be the expanded balls if they are chosen close enough in size to the originals. Of course small enlargements may require the use of lots of balls to cover $\mathcal{M}$.

4.5 Which Metric?

In order to make use of the preceding to construct a control cover of $\mathcal{M}$, we need a metric which at least guarantees that if κ stabilizes ν, {i.e., if $1+\kappa\nu$ has all its zeros in the open left-half plane}, then κ also stabilizes any transfer function in the ball $\mathbb{B}(\nu, r)$ for r sufficiently small. Picking a metric with this *robust stabilization property* is not altogether trivial. For example, although for sufficiently small g the transfer function

$$\tau_g = \frac{s-1+g}{(s+1)(s-1)}$$

can be made arbitrarily close to the transfer function

$$\nu = \tau_g\big|_{g=0} = \frac{1}{(s+1)}$$

in the metric space of normed differences between transfer function coefficients, for any controller transfer function κ which stabilizes ν one can always find a non-zero value of g sufficiently small such that κ does not stabilize τ_g at this value. This metric clearly does not have the property we seek. Two metrics which do are the gap metric [16] and the v-metric [17]. Moreover the v-metric is known to satisfy the continuity assumption stated just above Lemma 4.2 [14]. Thus we are able to construct a controller cover as follows.

4.6 Construction of a Control Cover

Suppose that the admissible process model transfer function class

$$\mathcal{M} = \bigcup_{p\in\mathcal{P}} \mathbb{B}(\nu_p, r_p)$$

is composed of balls $\mathbb{B}(\nu_p, r_p)$ which are open neighborhoods in the metric space of transfer functions with the v-metric μ. Suppose that these balls are each small enough so that for some sufficiently small continuous, positive function $p \longmapsto \rho_p$ we can construct for each $p \in \mathcal{P}$, a controller transfer function κ_p which stabilizes each transfer function in $\mathbb{B}(\nu_p, r_p+\rho_p)$. By Lemma 4.2, we can then construct a finite subset $\mathcal{Q} \subset \mathcal{P}$ such that

$$\mathcal{M} \subset \bigcup_{p\in\mathcal{Q}} \mathbb{B}(\nu_p, r_p + \rho_p)$$

By construction, κ_p stabilizes each transfer function in $\mathbb{B}(\nu_p, r_p+\rho_p)$. Thus for each $\tau \in \mathcal{M}$ there must be at least one value of $q \in \mathcal{Q}$ such that κ_q stabilizes τ. Since $\mathcal{Q}$ is a finite set, $\{\kappa_q : q \in \mathcal{Q}\}$ is a controller cover of $\mathcal{M}$.

5 Supervisory Control

Much has happened in adaptive control in the last forty years. The solution to the classical model reference problem is by now very well understood. Provably correct algorithms exist which, at least in theory, are capable of dealing with un-modelled dynamics, noise, right-half-plane zeros, and even certain types of nonlinearities. However despite these impressive gains, there remain many important, unanswered questions: Why, for example, is it still so difficult to explain to a novice why a particular algorithm is able to functions correctly in the face of un-modelled process dynamics and $\mathcal{L}^\infty$ bounded noise? How much un-modelled dynamics can a given algorithm tolerate before loop-stability is lost? How do we choose an adaptive control algorithm's many design parameters to achieve good disturbance rejection, transient response, etc.?

There is no doubt that there will eventually be satisfactory answers to all of these questions, that adaptive control will become much more accessible to non-specialists, that we will be able to much more clearly and concisely quantify un-modelled dynamics norm bounds, disturbance-to-controlled output gains, and so on and that because of this we will see the emergence of a bona fide computer-aided adaptive control design methodology which relies much more on design principals then on trial and error techniques. The aim of this chapter is to take a step towards these ends.

The intent of the chapter is to provide a relatively uncluttered analysis of the behavior of a set-point control system consisting of a poorly modelled process, an integrator and a multi-controller supervised by an estimator-based algorithm employing dwell-time switching. The system has been considered previously in [18, 19] where many of the ideas which follow were first presented. Similar systems have been analyzed in one form or another in [20, 21, 22] and elsewhere under various assumptions. It has been shown in [19] that the system's supervisor can successfully orchestrate the switching of a sequence of candidate set-point controllers into feedback with the system's imprecisely modelled siso process so as (i) to cause the output of the process to approach and track a constant reference input despite norm-bounded un-modelled dynamics, and constant disturbances and (ii) to insure that none of the signals within the overall system can grow without bound in response to bounded disturbances, be they constant or not. The objective of this chapter is to derive the same results in a much more straight forward manner. In fact this has already been done in [23] and [24] for a supervisory control system in which the switching between candidate controllers is constrained to be "slow." This restriction not only greatly simplified the analysis in comparison with that given in [19], but also made it possible to derive reasonably explicit upper bounds for the process's allowable un-modelled dynamics as well as for the system's disturbance-to-tracking error gain. In these notes we also constrain switching to be slow.

Adaptive set-point control systems typically consist of at least a process to be controlled, an integrator, a parameter tunable controller or "multi-

controller", a parameter estimator or "multi-estimator," and a tuner or "switching logic." In sharp contrast with non-adaptive linear control systems where subsystems are typically analyzed together using one overall linear model, in the adaptive case the sub-systems are not all linear and cannot be easily analyzed as one big inter-connected non-linear system. As a result, one needs to keep track of lots of equations, which can be quite daunting. One way to make things easier is to use carefully defined block diagrams which summarize equations and relations between signals in a way no set of equations can match. We make extensive use of such diagrams in this chapter.

5.1 The System

This section describes the overall structure of the supervisory control system to be considered. We begin with a description of the process.

Process = $\mathbb{P}$

The problem of interest is to construct a control system capable of driving to and holding at a prescribed set-point r, the output of a process modelled by a dynamical system with 'large' uncertainty. The process $\mathbb{P}$ is presumed to admit the model of a siso linear system whose transfer function from control input u to measured output y is a member of a continuously parameterized class of admissible transfer functions of the form

$$\mathcal{M} = \bigcup_{p \in \mathcal{P}} \{\nu_p + \delta : \|\delta\| \leq \epsilon_p\}$$

where $\mathcal{P}$ is a compact subset of a finite dimensional space,

$$\nu_p \stackrel{\Delta}{=} \frac{\alpha_p}{\beta_p}$$

is a pre-specified, strictly proper, *nominal transfer function*, ϵ_p is a real non-negative number, δ is a proper stable transfer function whose poles all have real parts less than the negative of a pre-specified *stability margin* $\lambda > 0$, and $\|\cdot\|$ is the shifted infinity norm

$$\|\delta\| = \sup_{\omega \in \mathbb{R}} |\delta(j\omega - \lambda)|$$

It is assumed that the coefficients of α_p and β_p depend continuously on p and for each $p \in \mathcal{P}$, that β_p is monic and that α_p and β_p are co-prime. All transfer functions in $\mathcal{M}$ are thus proper, but not necessarily stable rational functions. Prompted by the requirements of set-point control, it is further assumed that the numerator of each transfer function in $\mathcal{M}$ is non-zero at $s = 0$. The specific model of the process to be controlled is shown in Figure 8.

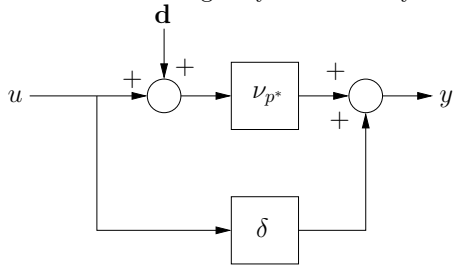

Fig. 8. Process Model

Here y is the process's measured output, $\mathbf{d}$ is a disturbance, and p^* is the index of the nominal process model transfer function which models $\mathbb{P}$.

We will consider the case when $\mathcal{P}$ is a finite set and also the case when $\mathcal{P}$ contains a continuum. Although both cases will be treated more or less simultaneously, there are two places where the continuum demands special consideration. The first is when one tries to construct stabilizable and detectable realizations of continuously parameterized transfer functions, which are continuous functions of p. In particular, unless the transfer functions in question all have the same McMillan degree, constructing realizations with all of these properties can be quite a challenge. The second place where the continuum requires special treatment, is when one seeks to characterize the key property implied by dwell time switching. Both of these matters will be addressed later in this chapter.

In the sequel, we define one by one, the component subsystems of the overall supervisory control system under consideration. We begin with the "multi-controller."

Multi-Controller $= \mathbb{C}_{\boldsymbol{\sigma}}$

We take as given a continuously parameterized family of "off-the-shelf" loop controller transfer functions $\mathcal{K} \stackrel{\Delta}{=} \{\kappa_p : p \in \mathcal{P}\}$ with at least the following property:

Stability Margin Property: *For each $p \in \mathcal{P}$, $-\lambda$ is greater than the real parts of all of the closed-loop poles*[3] *of the feedback interconnection*

We emphasize that stability margin property is only a minimal requirement on the κ_p. Actual design of the κ_p could be carried out using any one of a number of techniques, including linear quadratic or $\mathcal{H}^\infty$ methods, parameter-varying techniques, pole placement, etc.

We will also take as given an integer $n_C \geq 0$ and a continuously parameterized family of n_C-dimensional realizations $\{A_C(p), b_C(p),\ f_C(p), g_C(p)\}$,

[3] By the closed-loop poles are meant the zeros of the polynomial $s\rho_p\beta_p + \gamma_p\alpha_p$, where $\frac{\alpha_p}{\beta_p}$ and $\frac{\gamma_p}{\rho_p}$ are the reduced transfer functions ν_p and κ_p respectively.

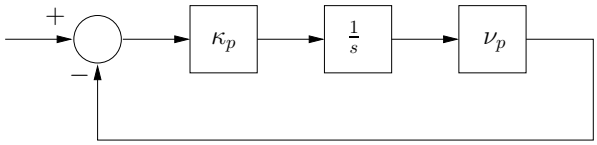

Fig. 9. Feedback Interconnection

one for each $\kappa_p \in \mathcal{K}$. These realizations are required to be chosen so that for each $p \in \mathcal{P}$, $(f_C(p), \lambda I + A_C(p))$ is detectable and $(\lambda I + A_C(p), b_C(p))$ is stabilizable. There are a great many different ways to construct such realizations, once one has in hand an upper bound n_κ on the McMillan Degrees of the κ_p. One is the $2n_\kappa$ - dimensional *identifier-based* realization

$$\left\{ \begin{pmatrix} A_I & 0 \\ 0 & A_I \end{pmatrix} + \begin{pmatrix} b_I \\ 0 \end{pmatrix} f_p,\ \begin{pmatrix} g_p b_I \\ b_I \end{pmatrix},\ f_p,\ g_p \right\} \tag{5.1}$$

where (A_I, b_I) is any given parameter-independent, n_κ-dimensional siso, controllable pair with A_I stable and f_p are g_p are respectively a parameter-dependent $1 \times 2n_\kappa$ matrix and a parameter dependent scalar. Another is the n_κ-dimensional *observer-based* realization

$$\{A_O + k_p f_O, b_p, f_O, g_p\}$$

where (f_O, A_O) is any given n_κ-dimensional, parameter-independent, observable pair with A_O stable and k_p and g_p are respectively a parameter-dependent $n_\kappa \times 1$ matrix and a parameter dependent scalar. Thus for the identifier-based realization $n_C = 2n_\kappa$ whereas for the observer-based realization $n_C = n_\kappa$. In either case, linear systems theory dictates that for a these realizations to exist, it is necessary to be able to represent each transfer function in $\mathcal{K}$ as a rational function $\rho(s)$ with a monic denominator of degree n_κ. Moreover, $\rho(s)$ must be defined so that the greatest common divisor of its numerator and denominator is a factor of the characteristic polynomial of A_I or A_O depending on which type of realization is being used. For the case when $\mathcal{K}$ is a finite set, this is easily accomplished by simply multiplying both the numerator and denominator of each reduced transfer function κ_p in $\mathcal{K}$ by an appropriate monic polynomial μ_p of sufficiently high degree so that the rational function $\rho(s)$ which results has a denominator of degree n_κ. Carry out this step for the case when $\mathcal{P}$ contains a continuum is more challenging because to obtain a realization which depends continuously on p, one must choose the coefficients of μ_p to depend continuously on p as well. One obvious way to side-step this problem is to deal only with the case when all of the transfer functions in $\mathcal{K}$ have the same McMillan degree, because in this case μ_p can always be chosen to be a fixed polynomial not depending on p. We will not pursue this issue in any greater depth in these notes. Instead, we will simply assume that such a continuously parameterized family of realizations of the κ_p has been constructed.

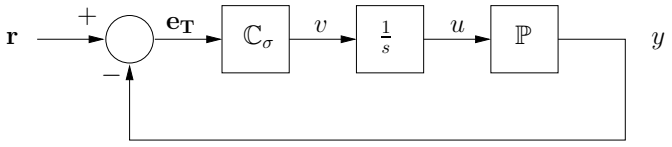

Fig. 10. Supervised Sub-System

Given such a family of realizations, the sub-system to be supervised is of the form shown in Figure 10
where $\mathbb{C}_\sigma$ is the n_C-dimensional switchable dynamical system

$$\dot{x}_C = A_C(\sigma)x_C + b_C(\sigma)\mathbf{e_T} \qquad v = f_C(\sigma)x_C + g_C(\sigma)\mathbf{e_T}, \tag{5.2}$$

called a *multi-controller*, v is the input to the *integrator*

$$\dot{u} = v, \tag{5.3}$$

$\mathbf{e_T}$ is the *tracking error*

$$\mathbf{e_T} \stackrel{\Delta}{=} r - y, \tag{5.4}$$

and σ is a piecewise constant *switching signal* taking values in $\mathcal{P}$.

Supervisor $= \mathbb{E} + \mathbb{W} + \mathbb{D}$

Our aim here is to define a "supervisor" which is capable of generating σ in real time so as to ensure both

1. *global boundedness* of all system signals in the face of an arbitrary but bounded disturbance inputs and
2. *set-point regulation* {i.e., $\mathbf{e_T} \to 0$} in the event that the disturbance signal is constant.

As we shall see, the kind of supervisor we will define will deliver even more – a form of exponential stability which will ensure that neither bounded measurement noise nor bounded system noise entering the system at any point can cause any signal in the system to grow without bound.

Parallel Realization of an Estimator-Based Supervisor

To understand the basic idea behind the type of supervisor we ultimately intend to discuss, it is helpful to first consider what we shall call a *parallel realized estimator-based supervisor*. This type of supervisor is applicable only when $\mathcal{P}$ is a finite set. So for the moment assume that $\mathcal{P}$ contains m elements $p_1, p_2, \ldots, p_m$ and consider the system shown in Figure 11.
Here each y_p is a suitably defined estimate of y which would be asymptotically correct if ν_p were the process model's transfer function and there were no noise or disturbances. The system which generates y_p would typically be an observer

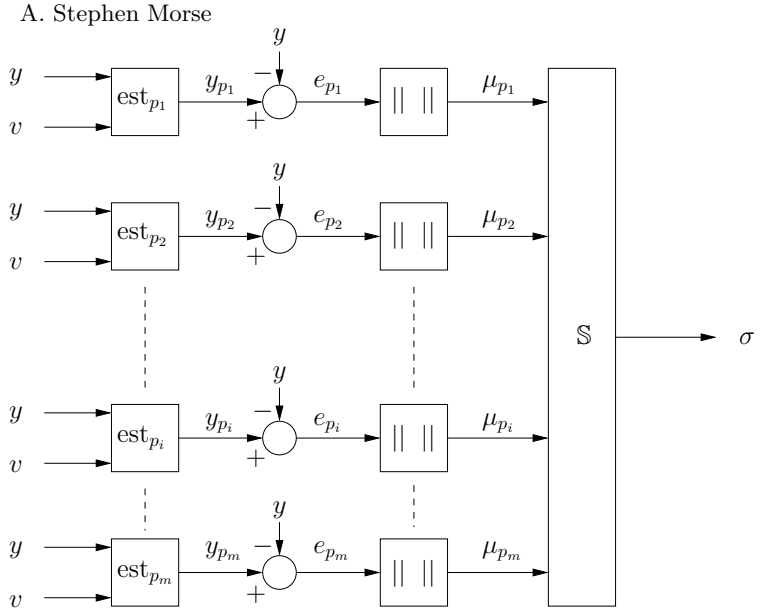

Fig. 11. Parallel Realization of an Estimator-Based Supervisor

for estimating the output of a linear realization of nominal process transfer function ν_p. For each $p \in \mathcal{P}$, $e_p = y_p - y$ denotes the pth output estimation error and μ_p is a suitably defined norm-squared value of e_p called a *monitoring signal* which is used by the supervisor to assess the potential performance of controller p. $\mathbb{S}$ is a switching logic whose function is to determine σ on the basis of the current values of the μ_p. The underlying decision making strategy used by such a supervisor is basically this: From time to time select for σ, that candidate control index q whose corresponding monitoring signal μ_q is the smallest among the μ_p, $p \in \mathcal{P}$. The justification for this heuristic idea is that the nominal process model whose associated monitoring signal is the smallest, "best" approximates what the process is and thus the candidate controller designed on the basis of this model ought to be able to do the best job of controlling the process.

Estimator-Based Supervisor

The supervisor shown in Figure 11 is a hybrid dynamical system whose inputs are v and y and whose output is σ. One shortcoming of this particular architecture is that it is only applicable when $\mathcal{P}$ is a finite set. The supervisor we will now define is functionally the same as the supervisor shown in Figure 11 but has a different realization, one which can be applied even when $\mathcal{P}$ contains a continuum of points. The supervisor consists of three subsystems: a *multi-estimator* $\mathbb{E}$, a *weight generator* $\mathbb{W}$, and a specific switching logic $\mathbb{S} \stackrel{\Delta}{=} \mathbb{D}$ called *dwell-time switching* .
We now describe each of these subsystems in greater detail.

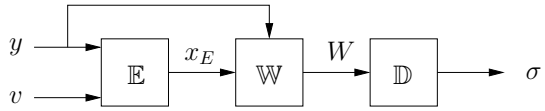

Fig. 12. Estimator-Based Supervisor

Multi-Estimator $= \mathbb{E}$

By a *multi-estimator* $\mathbb{E}$ for $\frac{1}{s}\mathcal{N}$, is meant an n_E-dimensional linear system

$$\dot{x}_E = A_E x_E + d_E y + b_E v \tag{5.5}$$

where

$$A_E = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}, \qquad d_E = \begin{pmatrix} b \\ 0 \end{pmatrix}, \qquad b_E = \begin{pmatrix} 0 \\ b \end{pmatrix},$$

Here (A, b) is any $(n_\nu + 1)$-dimensional, single input controllable pair chosen so that $(\lambda I + A)$ is exponentially stable, and n_ν is an upper bound on the McMillan degrees of the $\nu_p,\ p \in \mathcal{P}$. Because of this particular choice of matrices, it is possible to construct for each $p \in \mathcal{P}$, a row vector $c_E(p)$ for which

$$\{A_E + d_E c_E(p), b_E, c_E(p)\}$$

realizes $\frac{1}{s}\nu_p$ and $(\lambda I + A_E + d_E c(p), b_E)$ is stabilizable. We will assume that such a $c_E(p)$ has been constructed and we will further assume that it depends continuously on p. We will not explain how to carry out such a construction here, even though the procedure is straight forward if $\mathcal{P}$ is a finite set or if all the transfer functions in $\mathcal{N}$ have the same McMillan degree.

The parameter-dependent row vector $c_E(p)$ is used in the definition of $\mathbb{W}$ which will be given in Section 5.1. With $c_E(p)$ in hand it is also possible to define *output estimation errors*

$$e_p \stackrel{\Delta}{=} c_E(p) x_E - y, \quad p \in \mathcal{P} \tag{5.6}$$

While these error signals are not actually generated by the supervisor, they play an important role in explaining how the supervisor functions. It should be mentioned that for the case when $\mathcal{P}$ contains a continuum, the same issues arise in defining the quadruple $\{A_E + d_E c_E(p), b_E, c_E(p)\}$ to realize the $\frac{1}{s}\nu_p$, as were raised earlier when we discussed the problem of realizing the κ_p using the identifier-based quadruple in (5.1). Like before, we will sidestep these issues by assuming for the case when $\mathcal{P}$ is not finite, that all nominal process model transfer functions have the same McMillan degree.

Weight Generator $= \mathbb{W}$

The supervisor's second subsystem, $\mathbb{W}$, is a causal dynamical system whose inputs are x_E and y and whose state and output W is a symmetric "weighting matrix" which takes values in a linear space $\mathcal{W}$ of symmetric matrices.

W together with a suitably defined *monitoring function* $M : \mathcal{W} \times \mathcal{P} \to \mathbb{R}$ determine a scalar-valued *monitoring signal* of the form

$$\mu_p \stackrel{\Delta}{=} M(W, p) \tag{5.7}$$

which is viewed by the supervisor as a measure of the expected performance of controller p. $\mathbb{W}$ and M are defined by

$$\dot{W} = -2\lambda W + \begin{pmatrix} x_E \\ y \end{pmatrix} \begin{pmatrix} x_E \\ y \end{pmatrix}', \tag{5.8}$$

and

$$M(W, p) = \big(c_E(p) \; -1\big)\, W \, \big(c_E(p) \; -1\big)' \tag{5.9}$$

respectively, where $W(0)$ may be chosen to be any matrix in $\mathcal{W}$. The definitions of $\mathbb{W}$ and M are prompted by the observation that if μ_p are given by (5.7), then

$$\dot{\mu}_p = -2\lambda\mu_p + e_p^2, \quad p \in \mathcal{P}$$

because of (5.6), (5.8) and (5.9). Note that this implies that

$$\mu_p(T) = e^{-2\lambda T}||e_p||_T^2 \; + \; e^{-2\lambda T} M(W(0), p), \qquad T \geq 0, \; p \in \mathcal{P}$$

where, for any piecewise-continuous signal $z : [0, \infty) \to \mathbb{R}^n$, and any time $T > 0$, $||z||_T$ is the *exponentially weighted 2-norm*

$$||z||_T \stackrel{\Delta}{=} \sqrt{\int_0^T e^{2\lambda t}|z(t)|^2 dt}$$

Thus if $W(0) = 0$, $\mu_p(t)$ is simply a scaled version of the square of the exponentially weighted 2-norm of e_p.

Dwell-time Switching Logic $= \mathbb{D}$

The supervisor's third subsystem, called a *dwell-time switching logic* $\mathbb{D}$, is a hybrid dynamical system whose input and output are W and σ respectively, and whose state is the ordered triple $\{X, \tau, \sigma\}$. Here X is a discrete-time matrix which takes on sampled values of W, and τ is a continuous-time variable called a *timing signal*. τ takes values in the closed interval $[0, \tau_D]$, where τ_D is a pre-specified positive number called a *dwell time*. Also assumed pre-specified is a *computation time* $\tau_C \leq \tau_D$ which bounds from above for any $X \in \mathcal{W}$, the time it would take a supervisor to compute a value $p \in \mathcal{P}$ which minimizes $M(X, p)$. Between "event times," τ is generated by a reset integrator according to the rule $\dot{\tau} = 1$. Event times occur when the value of τ reaches either $\tau_D - \tau_C$ or τ_D; at such times τ is reset to either 0 or $\tau_D - \tau_C$ depending on the value of $\mathbb{D}$'s state. $\mathbb{D}$'s internal logic is defined by the flow diagram shown in Figure 13 where p_x denotes a value of $p \in \mathcal{P}$ which minimizes $M(X, p)$.

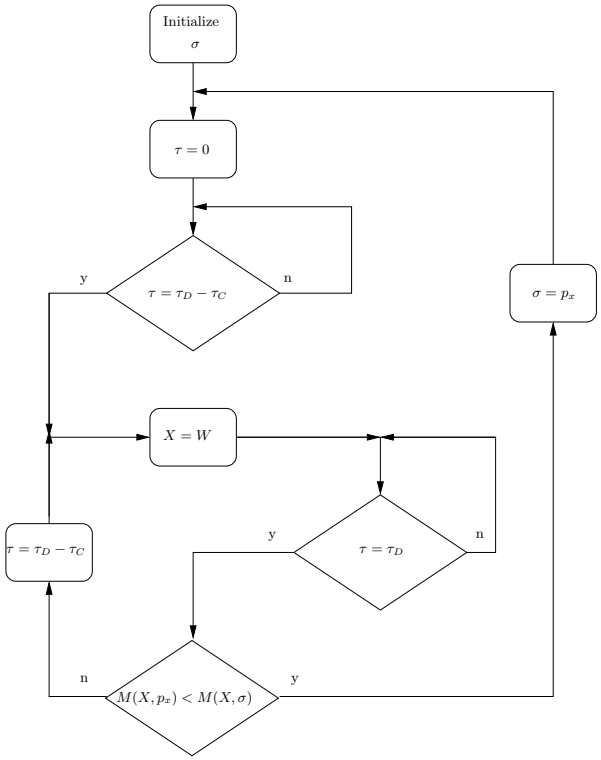

Fig. 13. Dwell-Time Switching Logic $\mathbb{D}$

Note that implementation of the supervisor just described can be accomplished when $\mathcal{P}$ contains either finitely or infinitely many points. However when $\mathcal{P}$ is a continuum, for the required minimization of $M(X, p)$ to be tractable, it will typically be necessary to make assumptions about both $c_E(p)$ and $\mathcal{P}$. For example, if $c_E(p)$ is an affine linear function and $\mathcal{P}$ is a finite union of convex sets, the minimization of $M(X, p)$ will be a finite family of finite-dimensional convex programming problems.

In the sequel we call a piecewise-constant signal $\bar{\sigma} : [0, \infty) \to \mathcal{P}$ *admissible* if it either switches values at most once, or if it switches more than once and the set of time differences between each two successive switching times is bounded below by τ_D. We write $\mathcal{S}$ for the set of all admissible switching signals. Because of the definition of $\mathbb{D}$, it is clear its output σ will be admissible. This means that switching cannot occur infinitely fast and thus that existence and uniqueness of solutions to the differential equations involved is not an issue.

Closed-Loop Supervisory Control System

The overall system just described, admits a block diagram description of the form shown in Figure 14. The basic properties of this system are summarized by the following theorem.

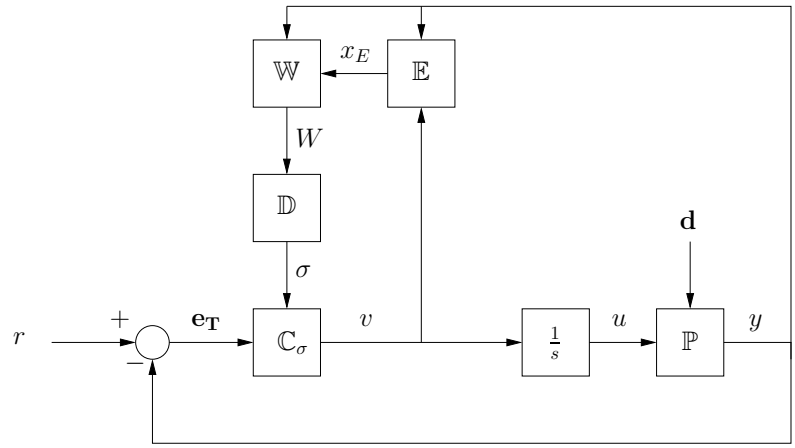

Fig. 14. Supervisory Control System

Theorem 5.1. *Let $\tau_C \geq 0$ be fixed. Let τ_D be any positive number no smaller than τ_C. There are positive numbers ϵ_p, $p \in \mathcal{P}$, for which the following statements are true provided the process $\mathbb{P}$ has a transfer function in $\mathcal{M}$.*

1. **Global Boundedness:** *For each constant set-point value r, each bounded piecewise-continuous disturbance input $\mathbf{d}$, and each system initialization, u, x_C, x_E, W, and X are bounded responses.*
2. **Tracking and Disturbance Rejection:** *For each constant set-point value r, each constant disturbance $\mathbf{d}$, and each system initialization, y tends to r and u, x_C, x_E, W, and X tend to finite limits, all as fast as $e^{-\lambda t}$.*

The theorem implies that the overall supervisory control system shown in Figure 14 has the basic properties one would expect of a non-adaptive linear set-point control system. It will soon become clear if it is not already that the induced $\mathcal{L}^2$ gain from $\mathbf{d}$ to $\mathbf{e_T}$ is finite as is the induced $\mathcal{L}^\infty$ gain from $\mathbf{d}$ to any state variable of the system.

5.2 Slow Switching

Although it is possible to establish correctness of the supervisory control system just described without any further qualification [19], in these notes we

will only consider the case when the switching between candidate controllers is constrained to be "slow" in a sense to be made precise below. This assumption not only greatly simplifies the analysis, but also make it possible to derive reasonably explicit bounds for the process's allowable un-modelled dynamics as well as for the system's disturbance-to-tracking-error gain.

Consider the system shown in Figure 15 which represents the feedback connection of linear systems with coefficient matrices $\{A_C(p), b_C(p), f_C(p), g_C(p)\}$ and $\{A_E + d_E c_E(p), b_E, c_E(p)\}$.

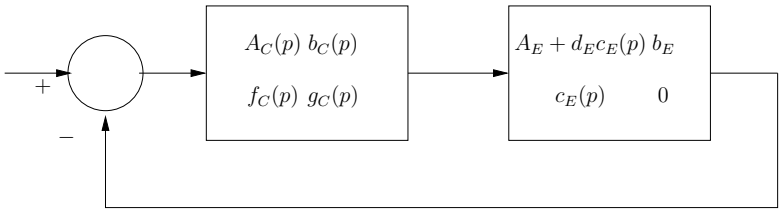

Fig. 15. Feedback Interconnection

With an appropriate ordering of substates, the "A" matrix for this system is

$$A_p = \begin{pmatrix} A_E + d_E(p)c_E(p) - b_E g_C(p)c_E(p) & b_E f_C(p) \\ -b_C(p)c_E(p) & A_C(p) \end{pmatrix} \tag{5.10}$$

Observe that the two subsystems shown in the figure are realizations of κ_p and $\frac{1}{s}\nu_p$ respectively. Thus because of the stability margin property discussed earlier, that factor of the characteristic polynomial of A_p determined by κ and $\frac{1}{s}\nu$ must have all its roots to the left of the vertical line $s = -\lambda$ in the complex plane. Any remaining eigenvalues of A_p must also line to the left of the line $s = -\lambda$ because $\{\lambda I + A_C(p), b_C(p), f_C(p), g_C(p)\}$ and $\{\lambda I + A_E + d_E c_E(p), b_E, c_E(p)\}$ are stabilizable and detectable systems. In other words, $\lambda I + A_p$ is an exponentially stable matrix and this is true for all $p \in \mathcal{P}$. In the sequel we assume the following.

Slow Switching Assumption: *Dwell time τ_D is large enough so that for each admissible switching signal $\sigma : [0, \infty) \to \mathcal{P}$, $\lambda I + A_\sigma$ is an exponentially stable matrix.*

Using Lemma 1.1 from Section 1.1, it is possible to compute an explicit lower bound for τ_D for which this assumption holds. Here's how. Since each $\lambda I + A_p$ is exponentially stable and $p \longmapsto A_p$ is continuous, it is possible to compute continuous, non-negative and positive functions $p \longmapsto t_p$ and $p \longmapsto \lambda_p$ respectively, such that

$$|e^{(\lambda I + A_p t)}| \leq e^{\lambda_p (t_p - t)}, \quad t \geq 0$$

It follows from Lemma 1.1 that if τ_D is chosen to satisfy

$$\tau_D > \sup_{p\in\mathcal{P}}\{t_p\}$$

then for each admissible switching signal σ, $\lambda I + A_\sigma$ will be exponentially stable.

5.3 Analysis

Our aim here is to establish a number of basic properties of the supervisory control system under consideration. We assume that $\mathbf{r}$ is an arbitrary but constant set-point value. In addition we invariably ignore initial condition dependent terms which decay to zero as fast as $e^{-\lambda t}$, as this will make things much easier to follow. A more thorough analysis which would take these terms into account can carried out in essentially the same manner.

Output Estimation Error e_{p^*}

Assume that the diagram in Figure 8 correctly models the process and consequently that p^* is the index of the correct nominal model transfer function ν_{p^*}. In this section we develop a useful formula for e_{p^*} where for $p \in \mathcal{P}$, e_p is the output estimation error

$$e_p = C_E x_E - y \tag{5.11}$$

defined previously by (5.6). In the sequel, for any signal w and polynomial $\alpha(s)$, we use the notation $\alpha(s)w$ to denote the action of the differential operator polynomial $\alpha(s)|_{s=\frac{d}{dt}}$ on w. For the sake of conciseness, we proceed formally, ignoring questions of differentiability of w. We will need the following easily verifiable fact.

Lemma 5.1. *For any triple of real matrices* $\{A_{n\times n},\ b_{n\times 1},\ c_{i\times n}\}$

$$c(sI - A)^{-1}b = \frac{\pi(s) - \bar{\pi}(s)}{\pi(s)}$$

where $\pi(s)$ *and* $\bar{\pi}(s)$ *are the characteristic polynomials of* A *and* $A + bc$ *respectively.*

A proof will not be given.

The process model depicted in Figure 8 implies that

$$\beta_{p^*} y = (\alpha_{p^*} + \beta_{p^*}\delta)u + \alpha_{p^*}\mathbf{d}$$

This and the fact that $\dot{u} = v$ enable us to write

$$s\beta_{p^*}y = (\alpha_{p^*} + \beta_{p^*}\delta)v + s\alpha_{p^*}\mathbf{d} \tag{5.12}$$

In view of (5.11)

$$e_{p^*} = c_E(p^*)x_E - y \tag{5.13}$$

Using this, is possible to re-write estimator equation $\dot{x}_E = A_E x_E + d_E y + b_E v$ defined by (5.5), as

$$\dot{x}_E = (A_E + d_E c_E(p^*))x_E - d_E e_{p^*} + b_E v \tag{5.14}$$

Since $\{A_E + d_E c_E(p^*), b_E, c_E(p^*)\}$ realizes $\frac{1}{s}\nu_{p^*}$ and $\nu_{p^*} = \frac{\alpha_{p^*}(s)}{\beta_{p^*}(s)}$ it must be true that

$$c_E(p^*)(sI - A_E - d_E c_E(p^*))^{-1}b_E = \frac{\alpha_{p^*}\theta(s)}{s\beta_{p^*}(s)\theta(s)}$$

where $s\beta_{p^*}(s)\theta(s)$ is the characteristic polynomial of $A_E + d_E c_E(p^*)$ and θ is a polynomial of unobservable-uncontrollable eigenvalues of $\{A_E + d_E c_E(p^*)$, $b_E, c_E(p^*)\}$. By assumption, $(s+\lambda)\theta$ is thus a stable polynomial. By Lemma 5.1,

$$c_E(p^*)(sI - A_E - d_E c_E(p^*))^{-1}d_E = \frac{\omega_E(s)\theta(s) - s\beta_{p^*}(s)\theta(s)}{s\beta_{p^*}(s)\theta(s)}$$

where $\omega_E(s)\theta(s)$ is the characteristic polynomial of A_E. These formulas and (5.14) imply that

$$s\beta_{p^*}\theta c_E(p^*)x_E = -(\omega_E\theta - s\beta_{p^*}\theta)e_{p^*} + \alpha_{p^*}\theta v$$

Therefore

$$s\beta_{p^*}c_E(p^*)x_E = -(\omega_E - s\beta_{p^*})e_{p^*} + \alpha_{p^*}v$$

This, (5.12) and (5.13) thus imply that

$$s\beta_{p^*}e_{p^*} = -(\omega_E - s\beta_{p^*})e_{p^*} - \beta_{p^*}\delta v - s\alpha_{p^*}\mathbf{d}$$

and consequently that

$$\omega_E e_{p^*} = -\beta_{p^*}\delta v - s\alpha_{p^*}\mathbf{d}$$

In summary, the assumption that $\mathbb{P}$ is modelled by the system shown in Figure 8, implies that the relationship between v, and e_{p^*} is as shown in Figure 16. Note that because of what has been assumed about δ and about the spectrum of A_E, the poles of all three transfer functions shown in this diagram lie to the left of the line $s = -\lambda$ in the complex plane.

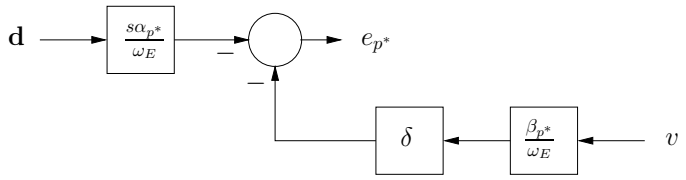

Fig. 16. Output Estimation Error e_{p^*}

Multi-Estimator/Multi-Controller Equations

Our next objective is to combine into a single model, the equations which describe $\mathbb{C}_\sigma$ and $\mathbb{E}$. As a first step, write $\bar{x}_E$ for the shifted state

$$\bar{x}_E = x_E + A_E^{-1} b_E r \tag{5.15}$$

Note that because of Lemma 5.1

$$c_E(p)(sI - A_E)^{-1} d_E = \frac{\omega(s) - s\beta_p(s)\theta_p(s)}{\omega(s)}, \quad p \in \mathcal{P}$$

where ω is the characteristic polynomial of A_E and for $p \in \mathcal{P}$, $s\beta_p\theta_p$ is the characteristic polynomial of $A_E + d_E c_E(p)$. Evaluation of this expression at $s = 0$ shows that

$$c_E(p) A_E^{-1} d_E = -1, \quad p \in \mathcal{P}$$

Therefore the pth output estimation error $e_p = c_E(p)x_E - y$ can be written as $e_p = c_E(p)\bar{x}_E + r - y$. But by definition, the tracking error is $e_{\mathbf{T}} = r - y$, so

$$e_p = c_E(p)\bar{x}_E + e_{\mathbf{T}}, \quad p \in \mathcal{P} \tag{5.16}$$

By evaluating this expression at $p = \sigma$, then solving for $e_{\mathbf{T}}$, one obtains

$$\mathbf{e_T} = e_\sigma - c_E(\sigma)\bar{x}_E \tag{5.17}$$

Substituting this expression for $\mathbf{e_T}$ into the multi-controller equations

$$\dot{x}_C = A_C(\sigma)x_C + b_C(\sigma)\mathbf{e_T} \qquad v = f_C(\sigma)x_C + g_C(\sigma)\mathbf{e_T}$$

defined by (5.2), yields

$$\dot{x}_C = A_C(\sigma)x_C - b_C(\sigma)c_E(\sigma)\bar{x}_E + b_C(\sigma)e_\sigma \qquad v = f_C(\sigma)x_C - g_C(\sigma)c_E(\sigma)\bar{x}_E + g_C(\sigma)e_\sigma \tag{5.18}$$

Note next that multi-estimator equation $\dot{x}_E = A_E x_E + d_E y + b_E v$ defined by (5.5) can be re-written using the shifted state $\bar{x}_E$ defined by (5.15) as

$$\dot{\bar{x}}_E = A_E\bar{x}_E - d_E(r - y) + b_E v$$

Therefore

$$\dot{\bar{x}}_E = A_E\bar{x}_E - d_E\mathbf{e_T} + b_E v$$

Substituting in the expression for $\mathbf{e_T}$ in(5.17) and the formula for v in (5.18) one gets

$$\dot{\bar{x}}_E = (A_E + d_E c_E(\sigma) - b_E g_C(\sigma)c_E(\sigma))\bar{x}_E + b_E f_C(\sigma)x_C + (b_E g_C(\sigma) - d_E)e_\sigma \tag{5.19}$$

Finally if we define the composite state

$$x = \begin{pmatrix} \bar{x}_E \\ x_C \end{pmatrix} \tag{5.20}$$

then it is possible to combine (5.18) and (5.19) into a single model

$$\dot{x} = A_\sigma x + b_\sigma e_\sigma \tag{5.21}$$

where for $p \in \mathcal{P}$, A_p is the matrix defined previously by (5.10) and

$$b_p = \begin{pmatrix} b_E g_C(p) - d_E \\ \\ b_C(p) \end{pmatrix}$$

The expressions for $\mathbf{e_T}$ and v in (5.17) and (5.18) can also be written in terms of x as

$$\mathbf{e_T} = e_\sigma + c_\sigma x \tag{5.22}$$

and

$$v = f_\sigma x + g_\sigma e_\sigma \tag{5.23}$$

respectively, where for $p \in \mathcal{P}$

$$c_p = -\begin{pmatrix} c_E(p) & 0 \end{pmatrix} \qquad f_p = \begin{pmatrix} -g_C(p)c_E(p) & f_C(p) \end{pmatrix} \qquad g_p = g_C(p)$$

Moreover, in view of (5.16),

$$e_p = e_q + c_{pq}x, \quad p, q \in \mathcal{P} \tag{5.24}$$

where

$$c_{pq} = \begin{pmatrix} c_E(p) - c_E(q) & 0 \end{pmatrix}, \quad p, q \in \mathcal{P} \tag{5.25}$$

Equations (5.20) - (5.24) can be thought of as an alternative description of $\mathbb{C}_\sigma$ and $\mathbb{E}$. We will make use of these equations a little later.

Exponentially Weighted 2-Norm

In section 5.1 we noted that each monitoring signal $\mu_p(t)$ could be written as

$$\mu_p(t) = e^{-2\lambda t}||e_p||_t^2 \; + \; e^{-2\lambda t}M(W(0), p), \qquad t \geq 0, \; p \in \mathcal{P}$$

where, for any piecewise-continuous signal $z : [0, \infty) \to \mathbb{R}^n$, and any time $t > 0$, $||z||_t$ is the *exponentially weighted 2-norm*

$$||z||_t \stackrel{\Delta}{=} \sqrt{\int_0^t e^{2\lambda\tau}|z(\tau)|^2 d\tau}$$

Since we are considering the case when $W(0) = 0$, the expression for μ simplifies to

$$\mu_p(t) = e^{-2\lambda t}||e_p||_t^2$$

Thus

$$||e_p||_t^2 = e^{2\lambda t}\mu_p(t), \quad p \in \mathcal{P} \tag{5.26}$$

The analysis which follows will be carried out using this exponentially weighted norm. In addition, for any time-varying SISO linear system Σ of the form $y = c(t)x + d(t)u$, $\dot{x} = A(t)x + b(t)u$ we write

$$\left\| \begin{matrix} A & b \\ c & d \end{matrix} \right\|$$

for the induced norm

$$\sup\{||y_u||_\infty : u \in \mathcal{U}\}$$

where y_u is Σ's zero initial state, output response to u and $\mathcal{U}$ is the space of all piecewise continuous signals u such that $||u||_\infty = 1$. The induced norm of Σ is finite whenever $\lambda I + A(t)$ is {uniformly} exponentially stable.

We note the following easily verifiable facts about the norm we are using. If $e^{-\lambda t}||u||_t$ is bounded on $[0, \infty)$ {in the $\mathcal{L}^\infty$ sense}, then so is y_u provided $d = 0$ and $\lambda I - A(t)$ is exponentially stable. If u is bounded on $[0, \infty)$ in the $\mathcal{L}^\infty$ sense, then so is $e^{-\lambda t}||u||_{\{0,t\}}$. If $u \to 0$ as $t \to \infty$, then so does $e^{-\lambda t}||u||_t$.

$\mathcal{P}$ is a Finite Set

It turns out that at this point the analysis for the case when $\mathcal{P}$ is a finite set, proceeds along a different path that the path to be followed in the case when $\mathcal{P}$ contains infinitely many points. In this section we focus exclusively on the case when $\mathcal{P}$ is finite.

As a first step, let us note that the relationships between $\mathbf{e_T}$, v and e_σ given by (5.22) - (5.24) can be conveniently represented by block diagrams which, in turn, can be added to the block diagram shown in Figure 16. What results in the block diagram shown in Figure 17.
In drawing this diagram we've set $p = \sigma$ and $q = p^*$ in (5.24) and we've represented the system defined by (5.21), (5.22) and (5.23) as two separate exponentially stable subsystems, namely

$$\begin{aligned} \dot{x}_1 &= A_\sigma x_1 + b_\sigma e_\sigma \\ v &= f_\sigma x_1 + g_\sigma e_\sigma \end{aligned} \qquad \begin{aligned} \dot{x}_2 &= A_\sigma x_2 + b_\sigma e_\sigma \\ \mathbf{e_T} &= c_\sigma x_2 + e_\sigma \end{aligned}$$

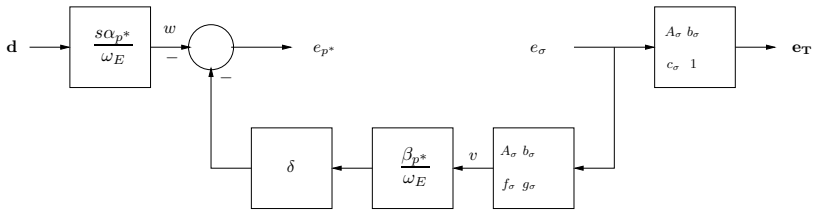

Fig. 17. A Representation of $\mathbb{P}$, $\mathbb{C}_\sigma$, $\mathbb{E}$ and $\frac{1}{s}$

where $x_1 = x_2 = x$. Note that the signal in the block diagram labelled w, will tend to zero if $\mathbf{d}$ is constant because of the zero at $s = 0$ in the numerator of the transfer function in the block driven by $\mathbf{d}$.

To proceed beyond this point, we will need to " close the loop" in the sense that we will need to relate e_σ to e_{p^*}. To accomplish this we need to address the consequences of dwell time switching, which up until now we've not considered.

Dwell-Time Switching

Note that each of the five blocks in Figure 17 represents an exponentially stable linear system with stability margin λ. Thus if it happened to be true that $e_\sigma = ge_{p^*}$ for some sufficiently small constant gain g then we would be able to deduce stability in the sense that the induced norm from d to $\mathbf{e_T}$ would be finite. Although no such gain exists, it nonetheless turns out to be true that there is a constant gain for which $||e_\sigma||_t \leq g||e_{p^*}||_t$ for all $t \geq 0$. To explain why this is so we will need the following proposition.

Proposition 5.1. *Suppose that $\mathcal{P}$ contains $m > 0$ elements that W is generated by (5.8), that the μ_p, $p \in \mathcal{P}$, are defined by (5.7) and (5.9), that $W(0) = 0$, and that σ is the response of $\mathbb{D}$ to W. Then for each time $T > 0$, there exists a piecewise constant function $\psi : [0, \infty) \to \{0, 1\}$ such that for all $q \in \mathcal{P}$,*

$$\int_0^\infty \psi(t)dt \leq m(\tau_D + \tau_C) \tag{5.27}$$

and

$$||(1 - \psi)e_\sigma + \psi e_q||_T \leq \sqrt{m}||e_q||_T \tag{5.28}$$

Proposition 5.1 highlights the essential consequences of dwell time switching needed to analyze the system under consideration for the case when $\mathcal{P}$ is finite. The proposition is proved in section 5.4.

A Snapshot at Time T

Fix $T > 0$ and let ψ be as in Proposition 5.1. In order to make use of (5.28), it is convenient to introduce the signal

$$z = (1 - \psi)e_\sigma + \psi e_{p^*} \tag{5.29}$$

since, with $q \stackrel{\Delta}{=} p^*$, (5.28) then becomes

$$||z||_T \leq \sqrt{m}||e_{p^*}||_T \tag{5.30}$$

Note that (5.24) implies that $e_\sigma = e_{p^*} + c_{\sigma p^*}x$. Because of this, the expression for z in (5.29) can be written as

$$z = (1 - \psi)e_\sigma + \psi(e_\sigma - c_{\sigma p^*}x)$$

Therefore after cancellation

$$z = e_\sigma - \psi c_{\sigma p^*}x$$

or

$$e_\sigma = \psi c_{\sigma p^*}x + z \tag{5.31}$$

Recall that

$$\dot{x} = A_\sigma x + b_\sigma e_\sigma \tag{5.32}$$

The point here is that (5.31) and (5.32) define a linear system with input z and output e_σ. We refer to this system as the *injected sub-system* of the overall supervisory control system under consideration. Adding a block diagram representation of this sub-system to the block diagram in Figure 17 results is the block diagram shown in Figure 18 which can be though of as a snapshot of the entire supervisory control system at time T. Of course the dashed block shown in the diagram is not really a block in the usual sense of signal flow. Nonetheless its inclusion in the diagram is handy for deriving norm bound inequalities, since in the sense of norms, the dashed block does provide the correct inequality, namely (5.30).

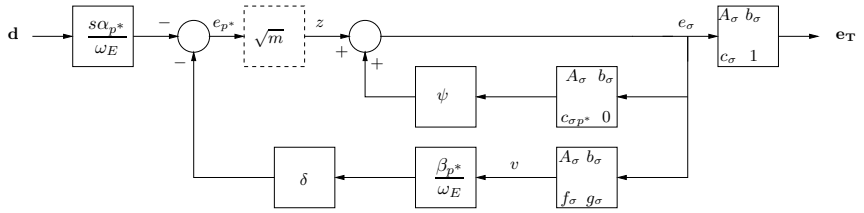

Fig. 18. A Snapshot of the Complete System at Time T

System Gains

Let us note that for any given admissible switching signal σ, each of the six blocks in Figure 18, excluding the dashed block and the block for ψ, represents an exponentially stable linear system with stability margin λ. It is convenient

at this point to introduce certain worst case "system gains" associated with these blocks. In particular, let us define for $p \in \mathcal{P}$

$$\boxed{\mathfrak{a}_p \triangleq \sqrt{2}\left|\frac{s\alpha_p}{\omega_E}\right|, \qquad \mathfrak{b}_p \triangleq \sqrt{2}\left|\frac{\beta_p}{\omega_E}\right| \sup_{\sigma\in\mathcal{S}} \left\|\begin{matrix} A_\sigma & b_\sigma \\ f_\sigma & g_\sigma \end{matrix}\right\|, \qquad \mathfrak{c} \triangleq \sup_{\sigma\in\mathcal{S}} \left\|\begin{matrix} A_\sigma & b_\sigma \\ c_\sigma & 1 \end{matrix}\right\|}$$

where, as defined earlier, $\|\cdot\|$ is the shifted infinity norm and $\mathcal{S}$ is the set of all admissible switching signals. In the light of Figure 18, it is easy to see that

$$||\mathbf{e_T}||_t \leq \mathfrak{c}||e_\sigma||_t, \quad t \geq 0, \tag{5.33}$$

and that

$$||e_{p^*}||_t \leq \epsilon_{p^*}\frac{\mathfrak{b}_{p^*}}{\sqrt{2}}|||e_\sigma||_t + \frac{\mathfrak{a}_{p^*}}{\sqrt{2}}||d||_t, \qquad t \geq 0, \tag{5.34}$$

where ϵ_{p^*} is the norm bound on δ.

To proceed we need an inequality which relates the norm of e_σ to the norm of z. For this purpose we introduce one more system gain, namely

$$\boxed{\mathfrak{v}_p \triangleq \sup_{\sigma\in\mathcal{S}}\sup_{t\geq 0}\int_0^t |c_{\sigma(t)p}\Phi(t,\tau)b_{\sigma(\tau)}e^{\lambda(t-\tau)}|^2 d\tau, \quad p \in \mathcal{P}}$$

where $\Phi(t,\tau)$ is the state transition matrix of A_σ. Note that each $\mathfrak{v}_p$ is finite because of the Slow Switching Assumption.

Analysis of the injected sub-system in Figure 18 can now be carried out as follows. Set

$$w_p(t,\tau) = c_{\sigma(t)p}\Phi(t,\tau)b_{\sigma(\tau)}$$

Using Cauchy-Schwartz

$$||\psi(w_{p^*} \circ e_\sigma)||_t \leq \sqrt{\mathfrak{v}_{p^*}\int_0^t \psi^2||e_\sigma||_\mu^2 d\mu}, \qquad t \geq 0 \tag{5.35}$$

where $w_{p^*} \circ e_\sigma$ is the zero initial state output response to e_σ of a system with weighting pattern w_{p^*}. From Figure 18 it is it is clear that $e_\sigma = z + \psi(w_{p^*} \circ e_\sigma)$. Thus taking norms

$$||e_\sigma||_t \leq ||z||_t + ||\psi(w_{p^*} \circ e_\sigma||_t$$

Therefore

$$||e_\sigma||_t^2 \leq 2||z||_t^2 + 2||\psi(w_{p^*} \circ e_\sigma||_t^2$$

Thus using (5.35)

$$||e_\sigma||_t^2 \leq 2||z||_t^2 + 2\mathfrak{v}_{p^*}\int_0^t \psi^2||e_\sigma||_\mu^2 d\mu, \quad 0 \leq t \leq T$$

Hence by the Bellman-Gronwall Lemma

$$||e_\sigma||_T^2 \le \left(2e^{2\mathfrak{v}_{p^*}\int_0^T \psi^2 dt}\right)||z||_T^2$$

so

$$||e_\sigma||_T \le \left(\sqrt{2}e^{\mathfrak{v}_{p^*}\int_0^T \psi^2 dt}\right)||z||_T$$

From this, (5.27), and the fact that $\psi^2 = \psi$, we arrive at

$$||e_\sigma||_T \le \left(\sqrt{2}e^{\mathfrak{v}_{p^*}m(\tau_D+\tau_C)}\right)||z||_T \tag{5.36}$$

Thus the induced gain from z to e_σ of the injected sub-system shown in Figure 18 is bounded above by $\sqrt{2}e^{\mathfrak{v}_{p^*}m(\tau_D+\tau_C)}$. We emphasize that this is a finite number, not depending on T.

Stability Margin

We've developed four key inequalities, namely (5.30), (5.33), (5.34) and (5.36) which we repeat below for ease of reference.

$$||z||_T \le \sqrt{m}||e_{p^*}||_T \tag{5.37}$$

$$||\mathbf{e_T}||_T \le \mathfrak{c}||e_\sigma||_T \tag{5.38}$$

$$||e_{p^*}||_T \le \epsilon_{p^*}\frac{\mathfrak{b}_{p^*}}{\sqrt{2}}|||e_\sigma||_T + \frac{\mathfrak{a}_{p^*}}{\sqrt{2}}||d||_T \tag{5.39}$$

$$||e_\sigma||_T \le \left(\sqrt{2}e^{\mathfrak{v}_{p^*}m(\tau_D+\tau_C)}\right)||z||_T \tag{5.40}$$

Inequalities (5.37), (5.39) and (5.40) imply that

$$||e_\sigma||_T \le \sqrt{m}e^{\mathfrak{v}_{p^*}m(\tau_D+\tau_C)})(\epsilon_{p^*}\mathfrak{b}_{p^*}||e_\sigma||_T + \mathfrak{a}_{p^*}||d||_T)$$

Thus if ϵ_{p^*} satisfies the *small gain condition*

$$\boxed{\epsilon_{p^*} < \frac{e^{-\mathfrak{v}_{p^*}m(\tau_D+\tau_C)}}{\mathfrak{b}_{p^*}\sqrt{m}}} \tag{5.41}$$

then

$$||e_\sigma||_T \le \frac{\mathfrak{a}_{p^*}}{\frac{e^{-\mathfrak{v}_{p^*}m(\tau_D+\tau_C)})}{\sqrt{m}} - \epsilon_{p^*}\mathfrak{b}_{p^*}}||\mathbf{d}||_T \tag{5.42}$$

The inequality in (5.41) provides an explicit upper bound for the norm of allowable unmodelled process dynamics, namely $|\!|\delta|\!|$.

A Bound on the Disturbance - to - Tracking - Error Gain

Note that (5.38) and (5.42) can be combined to provide an inequality of the form

$$||\mathbf{e_T}||_T \leq \mathfrak{g}_{p^*}||\mathbf{d}||_T, \quad T \geq 0 \tag{5.43}$$

where

$$\boxed{\mathfrak{g}_{p^*} = \frac{\mathfrak{ca}_{p^*}}{\frac{e^{-\mathfrak{v}_{p^*} m(\tau_D+\tau_C)}}{\sqrt{m}} - \epsilon_{p^*}\mathfrak{b}_{p^*}}} \tag{5.44}$$

The key point here is that $\mathfrak{g}_{p^*}$ does *not* depend on T even though the block diagram in Figure 18 does. Because of this, (5.43) must hold for all T. In other words, even though we've carried out an analysis at a fixed time T and in the process have had to define a several signals {e.g., ψ} which depended on T, in the end we've obtained an inequality namely (5.43), which is valid for all T. Because of this we can conclude that

$$||\mathbf{e_T}||_\infty \leq \mathfrak{g}_{p^*}||\mathbf{d}||_\infty \tag{5.45}$$

Thus $\mathfrak{g}_{p^*}$ bounds from above the overall disturbance - to - tracking - error gain of the system we've been studying.

Global Boundedness

The global boundedness condition of Theorem 5.1 can now easily be justified as follows. Suppose $\mathbf{d}$ is bounded on $[0,\ \infty)$ in the $\mathcal{L}^\infty$ sense. Then so must be $e^{-\lambda t}||\mathbf{d}||_t$. Hence by (5.42), $e^{-\lambda t}||e_\sigma||_t$ must be bounded on $[0,\ \infty)$ as well. This, the differential equation for x in (5.32), and the exponential stability of $\lambda I + A_\sigma$ then imply that x is also bounded on $[0,\ \infty)$. In view of (5.20) and (5.15), x_E and x_C must also be bounded. Next recall that the zeros of ω_E {i.e., the eigenvalues of A_E} have negative real parts less than $-\lambda$, and that the transfer function $\frac{\beta_{p^*}}{\omega_E}\delta$ in Figure 17 is strictly proper. From these observations, the fact that $e^{-\lambda t}||e_\sigma||_t$ is bounded on $[0,\infty)$, and the block diagram in Figure 17 one readily concludes that e_{p^*} is bounded on $[0,\infty)$. From (5.24), $e_\sigma = e_{p^*} + c_{\sigma p^*}x$. Therefore e_σ is bounded on $[0,\infty)$. Boundedness of $\mathbf{e_T}$ and v follow at once from (5.22) and (5.23) respectively. In view of (5.4), y must be bounded. Thus W must be bounded because of (5.8). Finally note that u must be bounded because of the boundedness of y and v and because of our standing assumption that the transfer function of $\mathbb{P}$ is non-zero at $s = 0$. This, in essence, proves Claim 1 of Theorem 5.1.

Convergence

Now suppose that $\mathbf{d}$ is a constant. Examination of Figure 17 reveals that w must tend to zero as fast as $e^{-\lambda t}$ because of the zero at $s = 0$ in the numerator

of the transfer function from $\mathbf{d}$ to w. Thus $||w||_\infty < \infty$. Figure 17 also implies that

$$||e_{p^*}||_T \leq ||w||_T + \epsilon_{p^*} \frac{b_{p^*}}{\sqrt{2}} ||e_\sigma||_T$$

This (5.37), (5.40) and (5.41) implies that

$$||e_\sigma||_T \leq \frac{\sqrt{2}}{\frac{e^{-\mathfrak{v}_{p^*} m(\tau_D+\tau_C)})}{\sqrt{m}} - \epsilon_{p^*} \mathfrak{b}_{p^*}} ||w||_T$$

Since this inequality holds for all $T \geq 0$, it must be true that $||e_\sigma||_\infty < \infty$. Hence e_σ must tend to zero as fast as $e^{-\lambda t}$. So therefore must x because of the differential equation for x in (5.21). In view of (5.20) $\bar{x}_E$ and x_C must tend to zero. Thus x_E must tend to $A_E^{-1} b_E r$ because of (5.15). Moreover e_{p^*} must tend to zero as can be plainly seen from Figure 17. Hence from the formulas (5.22) and(5.23) for $\mathbf{e_T}$ and v respectively one concludes that these signals must tend to zero as well. In view of (5.4), y must tent to r . Thus W must approach a finite limit because of (5.8). Finally note that u tend to a finite limit because y and v do and because of our standing assumption that the transfer function of $\mathbb{P}$ is non-zero at $s = 0$. This, in essence, proves Claim 2 of Theorem 5.1.

$\mathcal{P}$ is not a finite set

We now consider the more general case when $\mathcal{P}$ is a compact but not necessarily finite subset of a finite dimensional linear space. The following proposition replaces Proposition 5.1 which is clearly not applicable to this case. The proposition relies on the fact that every nominal transfer function in $\mathcal{N}$ can be modelled by a linear system of dimension at most n_E.

Dwell-Time Switching

Proposition 5.2. *Suppose that $\mathcal{P}$ is a compact subset of a finite dimensional space, that $p \longmapsto c_E(p)$ is a continuous function taking values in $\mathbb{R}^{1\times n_E}$, that $c_{pq} = \big(c_E(p) - c_E(q)\ 0\big)$ as in (5.25), that W is generated by (5.8), that the μ_p, $p \in \mathcal{P}$, are defined by (5.7) and (5.9), that $W(0) = 0$, and that σ is the response of $\mathbb{D}$ to W. For each $q \in \mathcal{P}$, each real number $\rho > 0$ and each fixed time $T > 0$, there exists piecewise-constant signals $h : [0, \infty) \to \mathbb{R}^{1\times(n_E+n_C)}$ and $\psi : [0, \infty) \to \{0, 1\}$ such that*

$$|h(t)| \leq \rho, \quad t \geq 0 \tag{5.46}$$

$$\int_0^\infty \psi(t)dt \leq n_E(\tau_D + \tau_C) \tag{5.47}$$

and

$$||(1-\psi)(e_\sigma - hx) + \psi e_q||_T \leq \left\{1 + 2n_E \left(\frac{1 + \sup_{p\in\mathcal{P}} |c_{pq}|}{\rho}\right)^{n_E}\right\} ||e_q||_T \tag{5.48}$$

This Proposition is proved in Section 5.4. Proposition 5.2 summarizes the key consequences of dwell time switching which are needed to analyze the system under consideration. The term involving h in (5.48) present some minor difficulties which we will deal with next.

Let us note that for any piece-wise continuous matrix-valued signal $h : [0,\infty) \to \mathbb{R}^{1\times(n_E+n_C)}$, it is possible to re-write the equations (5.21) - (5.22) as

$$\dot{x} = (A_\sigma + b_\sigma h)x + b_\sigma \bar{e} \tag{5.49}$$

$$\mathbf{e_T} = \bar{e} + (c_\sigma + h)x \tag{5.50}$$

and

$$v = (f_\sigma + g_\sigma h)x + g_\sigma \bar{e} \tag{5.51}$$

respectively where

$$\bar{e} = e_\sigma - hx \tag{5.52}$$

Note also that the matrix $\lambda I + A_\sigma + b_\sigma h$ will be exponentially stable for any $\sigma \in \mathcal{S}$ if $|h| \le \rho, t \ge 0$, where ρ is a sufficiently small positive number. Such a value of ρ exists because $p \longmapsto A_p$ and $p \longmapsto b_p$ are continuous and bounded functions on $\mathcal{P}$ and because $\lambda I + A_\sigma$ is exponentially stable for every admissible switching signal. In the sequel we will assume that ρ is such a number and that $\mathcal{H}$ is the set of all piece-wise continuous signals h satisfying $|h| \le \rho, t \ge 0$.

A Snapshot at Time T

Now fix $T > 0$ and let ψ and h be signals for which (5.46) - (5.48) hold with $q = p^*$. To account for any given h in (5.49) - (5.51), we will use in place of the diagram in Figure 17, the diagram shown in Figure 19. As with the representation in Figure 17, we are representing the system defined by (5.49) - (5.51) as two separate subsystems, namely

$$\begin{aligned} \dot{x}_1 &= (A_\sigma + b_\sigma h)x_1 + b_\sigma \bar{e} \\ v &= (f_\sigma + g_\sigma h)x_1 + g_\sigma \bar{e} \end{aligned} \qquad \begin{aligned} \dot{x}_2 &= (A_\sigma + b_\sigma h)x_2 + b_\sigma \bar{e} \\ \mathbf{e_T} &= (c_\sigma + h)x_2 + \bar{e} \end{aligned}$$

where $x_1 = x_2 = x$.

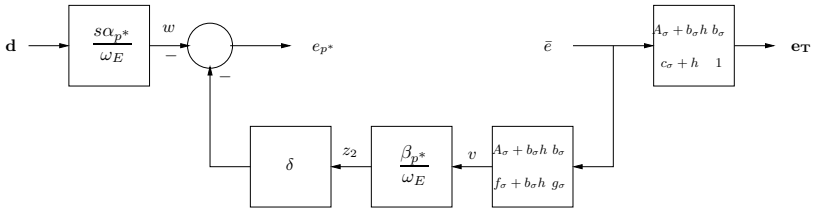

Fig. 19. A Representation of $\mathbb{P}$, $\mathbb{C}_\sigma$, $\mathbb{E}$ and $\frac{1}{s}$

Note that each of the five blocks in Figure 19 represents an exponentially stable linear system with stability margin λ.

In order to make use of (5.48), it is helpful to introduce the signal

$$\bar{z} = (1 - \psi)(e_\sigma - hx) + \psi e_{p^*} \tag{5.53}$$

since (5.48) then becomes

$$||\bar{z}||_T \leq \gamma_{p^*} ||e_{p^*}||_T \tag{5.54}$$

where

$$\gamma_{p^*} \triangleq \left\{ 1 + 2n_E \left(\frac{1 + \sup_{p \in \mathcal{P}} |c_{pp^*}|}{\rho} \right)^{n_E} \right\}$$

Note that

$$e_\sigma = e_{p^*} + c_{\sigma p^*} x$$

because of (5.24). Solving for e_{p^*} in substituting the result into (5.53) gives

$$\bar{z} = (1 - \psi)(e_\sigma - hx) + \psi(e_\sigma - c_{\sigma p^*} x)$$

Thus

$$\bar{z} = e_\sigma - hx - \psi(c_{\sigma p^*} - h)x$$

In view of (5.52) we can therefore write

$$\bar{z} = \bar{e} - \psi(c_{\sigma p^*} - h)x$$

or

$$\bar{e} = \bar{z} + \psi(c_{\sigma p^*} - h)x \tag{5.55}$$

Recall that (5.49) states that

$$\dot{x} = (A_\sigma + b_\sigma h)x + b_\sigma \bar{e} \tag{5.56}$$

Observe that (5.55) and (5.56) define a linear system with input $\bar{z}$ and output $\bar{e}$ which we refer to as the *injected system* for the problem under consideration. Adding a block diagram representation of this sub-system to the block diagram in Figure 19 results in the block diagram shown in Figure 20. Just as in the case when $\mathcal{P}$ is finite, the dashed block is not really a block in the sense of signal flow.

System Gains

Let us note that for any given admissible switching signal σ, each of the six blocks in Figure 20, excluding the dashed block and the block for ψ, represents an exponentially stable linear system with stability margin λ. It is convenient at this point to introduce certain worst case "system gains" associated with these blocks. In particular, let us define for $p \in \mathcal{P}$

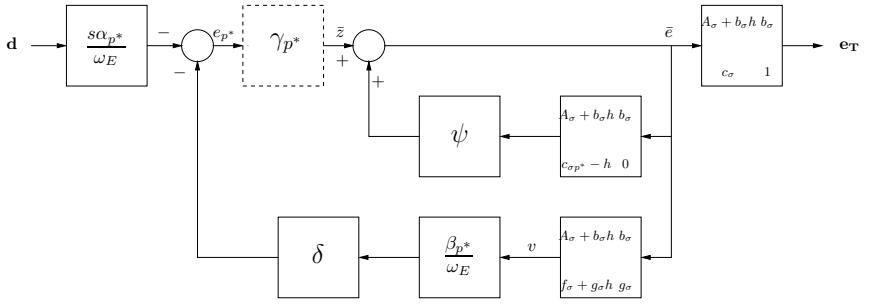

Fig. 20. A Snapshot of the Complete System at Time T

$$\bar{\mathfrak{b}}_p \triangleq \sqrt{2}\left|\frac{\beta_p}{\omega_E}\right| \left\{ \sup_{h\in\mathcal{H}} \sup_{\sigma\in\mathcal{S}} \left\| \begin{matrix} A_\sigma + b_\sigma h & b_\sigma \\ f_\sigma + g_\sigma h & g_\sigma \end{matrix} \right\| \right\} \qquad \bar{\mathfrak{c}} \triangleq \sup_{h\in\mathcal{H}} \sup_{\sigma\in\mathcal{S}} \left\| \begin{matrix} A_\sigma + b_\sigma h & b_\sigma \\ c_\sigma + h & 1 \end{matrix} \right\|$$

In the light of Figure 20, it is easy to see that

$$||\mathbf{e_T}||_t \leq \bar{\mathfrak{c}}||\bar{e}||_t, \quad t \geq 0, \tag{5.57}$$

and that

$$||e_{p^*}||_t \leq \epsilon_{p^*} \frac{\bar{\mathfrak{b}}_{p^*}}{\sqrt{2}}||\bar{e}||_t + \frac{\mathfrak{a}_{p^*}}{\sqrt{2}}||d||_t, \qquad t \geq 0, \tag{5.58}$$

where ϵ_{p^*} is the norm bound on δ and $\mathfrak{a}_p$ is as defined in section 5.3.

To proceed we need an inequality which related the norm of $\bar{e}$ to the norm of $\bar{z}$. For this purpose we introduce the additional system gain

$$\bar{\mathfrak{v}}_q \triangleq \sup_{h\in\mathcal{H}} \sup_{\sigma\in\mathcal{S}} \sup_{t\geq 0} \int_0^t |(c_{\sigma(t)q} - h(t))\Phi(t,\tau)b_{\sigma(\tau)}e^{\lambda(t-\tau)}|^2 d\tau$$

where $\Phi(t,\tau)$ is the state transition matrix of $A_\sigma + b_\sigma h$. Note that each $\bar{\mathfrak{v}}_q$ is finite because of the Slow Switching Assumption.

Analysis of the injected sub-system shown in Figure 20 is the same as in the case when $\mathcal{P}$ is finite. Instead of (5.36), what one obtains in this case is the inequality

$$||\bar{e}||_T \leq \left(\sqrt{2}e^{\bar{\mathfrak{v}}_{p^*} n_E(\tau_D+\tau_C)}\right)||\bar{z}||_T \tag{5.59}$$

Stability Margin

We've developed four key inequalities for the problem at hand, namely (5.54), (5.57), (5.58) and (5.59) which we repeat for ease of reference.

$$||\bar{z}||_T \leq \gamma_{p^*}||e_{p^*}||_T \tag{5.60}$$

$$||\mathbf{e_T}||_T \leq \bar{\mathfrak{c}}||\bar{e}||_T \tag{5.61}$$

$$||e_{p^*}||_T \leq \epsilon_{p^*}\frac{\bar{\mathfrak{b}}_{p^*}}{\sqrt{2}}|||\bar{e}||_T + \frac{\mathfrak{a}_{p^*}}{\sqrt{2}}||d||_T \tag{5.62}$$

$$||\bar{e}||_T \leq \left(\sqrt{2}e^{\bar{\mathfrak{v}}_{p^*}m(\tau_D+\tau_C)}\right)||\bar{z}||_T \tag{5.63}$$

Observe that except for different symbols, these inequalities are exactly the same as those in (5.37) - (5.40) respectively. Because of this, we can state at once that if ϵ_{p^*} satisfies the small gain condition

$$\boxed{\epsilon_{p^*} < \frac{e^{-\bar{\mathfrak{v}}_{p^*}n_E(\tau_D+\tau_C)}}{\bar{\mathfrak{b}}_{p^*}\gamma_{p^*}}} \tag{5.64}$$

then

$$||\bar{e}||_T \leq \frac{\mathfrak{a}_{p^*}}{\frac{e^{-\bar{\mathfrak{v}}_{p^*}n_E(\tau_D+\tau_C)}}{\gamma_{P^*}} - \epsilon_{p^*}\bar{\mathfrak{b}}_{p^*}}||\mathbf{d}||_T, \tag{5.65}$$

As in the case when $\mathcal{P}$ is finite, inequality in (5.64) provides an explicit bound for the norm of allowable process dynamics.

A Bound on the Disturbance - to - Tracking - Error Gain

Note that (5.61) and (5.65) can be combined to provide an inequality of the form

$$||\mathbf{e_T}||_T \leq \bar{\mathfrak{g}}_{p^*}||\mathbf{d}||_T, \quad T \geq 0$$

where

$$\boxed{\bar{\mathfrak{g}}_{p^*} = \frac{\bar{\mathfrak{c}}\mathfrak{a}_{p^*}}{\frac{e^{-\bar{\mathfrak{v}}_{p^*}n_E(\tau_D+\tau_C)}}{\gamma_{P^*}} - \epsilon_{p^*}\bar{\mathfrak{b}}_{p^*}}} \tag{5.66}$$

Moreover, because the preceding inequality holds for all $T > 0$ and $\mathfrak{g}_{p^*}$ is independent of T, it must be true that

$$||\mathbf{e_T}||_\infty \leq \bar{\mathfrak{g}}_{p^*}||\mathbf{d}||_\infty$$

Thus for the case when $\mathcal{P}$ contains infinitely many points, $\bar{\mathfrak{g}}_{p^*}$ bounds from above the overall system's disturbance - to - tracking - error gain.

Global Boundedness and Convergence

It is clear from the preceding that the reasoning for the case when $\mathcal{P}$ contains infinitely many points parallels more or less exactly the reasoning used for the case when $\mathcal{P}$ contains only finitely many points. Thus for example, the claims of Theorem 5.1 regarding global boundedness and exponential convergence for the case when $\mathcal{P}$ contains infinitely many points, can be established in essentially the same way as which they were established earlier in these notes for the case when $\mathcal{P}$ is a finite set. For this reason, global boundedness and convergence arguments will not be given here.

5.4 Analysis of the Dwell Time Switching Logic

We now turn to the analysis of dwell time switching. In the sequel, $T > 0$ is fixed, σ is a given switching signal, $t_0 \stackrel{\Delta}{=} 0$, t_i denotes the ith time at which σ switches and p_i is the value of σ on $[t_{i-1}, t_i)$; if σ switches at most $n < \infty$ times then $t_{n+1} \stackrel{\Delta}{=} \infty$ and p_{n+1} denotes σ's value on $[t_n, \infty)$. Any time X takes on the current value of W is called a *sample time*. We use the notation $\lfloor t \rfloor$ to denote the sample time just preceding time t, if $t > \tau_D - \tau_C$, and the number zero otherwise. Thus, for example, $\lfloor t_0 \rfloor = 0$ and $\lfloor t_i \rfloor = t_i - \tau_C,\ i > 0$. We write k for that integer for which $T \in [t_{k-1}, t_k)$. For each $j \in \{1, 2, \ldots, k\}$ define

$$\bar{t}_j = \begin{cases} t_j & \text{if } j < k \\ T & \text{if } j = k \end{cases},$$

and let $\phi_j : [0, \infty) \to \{0, 1\}$ be that piecewise-constant signal which is zero everywhere except on the interval

$$[\lfloor t_j \rfloor,\ \bar{t}_j), \qquad \text{if } \ \bar{t}_j - t_{j-1} \leq \tau_D$$

or

$$[\lfloor \bar{t}_j \rfloor - \tau_C,\ \bar{t}_j), \ \text{ if } \ \bar{t}_j - t_{j-1} > \tau_D$$

In either case ϕ_j has support no greater than $\tau_D + \tau_C$ and is idempotent {i.e., $\phi_j^2 = \phi_j$}. The following lemma describes the crucial consequence of dwell time switching upon which the proofs of Proposition 5.1 and 5.2 depend.

Lemma 5.2. *For each* $j \in \{1, 2, \ldots, k\}$

$$||(1 - \phi_j)e_{p_j}||_{\bar{t}_j} \leq ||(1 - \phi_j)e_q||_T, \qquad \forall q \in \mathcal{P}$$

Proof of Lemma 5.2: The definition of dwell time switching implies that

$$\mu_{p_j}(\lfloor t_{j-1} \rfloor) \leq \mu_q(\lfloor t_{j-1} \rfloor), \qquad \forall q \in \mathcal{P},$$

$$\mu_{p_j}(\lfloor \bar{t}_j \rfloor - \tau_C) \leq \mu_q(\lfloor \bar{t}_j \rfloor - \tau_C), \forall q \in \mathcal{P} \ \text{ if } \bar{t}_j - t_{j-1} > \tau_D$$

As noted earlier, for all $t \geq 0$

$$\mu_p(t) = e^{-2\lambda t}||e_p||_t^2\ ,\ p \in \mathcal{P}$$

Therefore

$$\left.\begin{aligned} ||e_{p_j}||^2_{\lfloor t_{j-1} \rfloor} &\leq ||e_q||^2_{\lfloor t_{j-1} \rfloor}, \qquad \forall q \in \mathcal{P} \\ ||e_{p_j}||^2_{(\lfloor \bar{t}_j \rfloor - \tau_C)} &\leq ||e_q||^2_{(\lfloor \bar{t}_j \rfloor - \tau_C)}, \forall q \in \mathcal{P}, \text{ if } \bar{t}_j - t_j > \tau_D \end{aligned}\right\} \tag{5.67}$$

The definitions of ϕ_j implies that for $l \in \mathcal{P}$

$$||(1-\phi_j)e_l||^2_{\bar{t}_j} = \begin{cases} ||(1-\phi_j)e_l||^2_{\lfloor t_{j-1}\rfloor} & \text{if } \bar{t}_j - t_{j-1} \le \tau_D \\ ||(1-\phi_j)e_l||^2_{(\lfloor \bar{t}_j\rfloor - \tau_C)} & \text{if } \bar{t}_j - t_{j-1} > \tau_D \end{cases}$$

From this and (5.67) we obtain for all $q \in \mathcal{P}$

$$||(1-\phi_j)e_{p_j}||^2_{\bar{t}_j} = ||e_{p_j}||^2_{\lfloor t_{j-1}\rfloor} \le ||e_q||^2_{\lfloor t_{j-1}\rfloor} \le ||e_q||^2_{\bar{t}_j} = ||(1-\phi_j)e_q||^2_{\bar{t}_j}$$

if $\bar{t}_j - t_{j-1} \le \tau_D$ and

$$||(1-\phi_j)e_{p_j}||^2_{\bar{t}_j} = ||e_{p_j}||^2_{\lfloor \bar{t}_j - \tau_C\rfloor} \le ||e_q||^2_{\lfloor \bar{t}_j - \tau_C\rfloor} \le ||e_q||^2_{\bar{t}_j} = ||(1-\phi_j)e_q||^2_{\bar{t}_j}$$

if $\bar{t}_j - t_{j-1} > \tau_D$. From this and the fact that

$$||(1-\phi_j)e_q||^2_{\bar{t}_j} \le ||(1-\phi_j)e_q||^2_T,\ q \in \mathcal{P},$$

there follows

$$||(1-\phi_j)e_{p_j}||_{\bar{t}_j} \le ||(1-\phi_j)e_q||_T, \quad \forall\ q \in \mathcal{P}$$

■

Implication of Dwell - Time Switching When $\mathcal{P}$ is a Finite Set

The proof of Proposition 5.1 makes use of the following lemma.

Lemma 5.3. *For all* $\mu_i \in [0,1],\ i \in \{1,2,\dots,m\}$

$$\sum_{i=1}^{m}(1-\mu_i) \le (m-1) + \prod_{i=1}^{m}(1-\mu_i) \tag{5.68}$$

Proof of Lemma 5.3: Set $x_i = 1-\mu_i,\ i \in \{1,2,\dots,\}$. It is enough to show that for $x_i \in [0,1],\ i \in \{1,2,\dots,\}$

$$\sum_{i=1}^{j} x_i \le (j-1) + \prod_{i=1}^{j} x_j \tag{5.69}$$

for $j \in \{1,2,\dots,m\}$. Clearly (5.69) is true if $j = 1$. Suppose therefore that for some $k > 0$, (5.69) holds for $j \in \{1,2,\dots,k\}$. Then

$$\begin{aligned} \sum_{i=1}^{k+1} x_i &= x_{k+1} + \sum_{i=1}^{k} x_i \\ &\le x_{k+1} + (k-1) + \prod_{i=1}^{k} x_i \\ &\le (1-x_{k+1})\left(1 - \prod_{i=1}^{k} x_i\right) + x_{k+1} + (k-1) + \prod_{i=1}^{k} x_i \\ &= k + \prod_{i=1}^{k+1} x_i \end{aligned}$$

By induction, (5.69) thus holds for $j \in \{1, 2, \ldots, m\}$. ■

Proof of Proposition 5.1: For each distinct $p \in \{p_1, p_2, \ldots, p_k\}$, let $\mathcal{I}_p$ denote the set of nonnegative integers i such that $p_i = p$ and write j_p for the largest integer in $\mathcal{I}_p$. Note that $j_{p_k} = k$. Let $\mathcal{J}$ denote the set of all such j. since $\mathcal{P}$ contains m elements, m bounds from above the number of elements in $\mathcal{J}$. For each $j \in \mathcal{J}$, define

$$\psi \stackrel{\Delta}{=} 1 - \prod_{j \in \mathcal{J}} (1 - \phi_j), \tag{5.70}$$

Since each ϕ_p has co-domain $\{0, 1\}$, support no greater than $\tau_D + \tau_C$ and is idempotent, it must be true that ψ has co-domain $\{0, 1\}$, support no greater than $m(\tau_D + \tau_C)$ and is idempotent as well. Therefore, (5.27) holds.

Now

$$||(1-\psi)e_\sigma||_T^2 = \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}_{p_j}} (||(1-\psi)e_{p_j}||_{\bar{t}_i}^2 - ||(1-\psi)e_{p_j}||_{t_{i-1}}^2) \leq \sum_{j \in \mathcal{J}} ||(1-\psi_)e_{p_j}||_{\bar{t}_j}^2 \tag{5.71}$$

In view of (5.70) we can write

$$\sum_{j \in \mathcal{J}} ||(1-\psi)e_{p_j}||_{\bar{t}_j}^2 = \sum_{j \in \mathcal{J}} \left\| \left\{ \prod_{l \in \mathcal{J}} (1 - \phi_l) \right\} e_{p_j} \right\|_{\bar{t}_j}^2 \tag{5.72}$$

But

$$\sum_{j \in \mathcal{J}} \left\| \left\{ \prod_{l \in \mathcal{J}} (1 - \phi_l) \right\} e_{p_j} \right\|_{\bar{t}_j}^2 \leq \sum_{j \in \mathcal{J}} \left\| (1 - \phi_j) e_{p_j} \right\|_{\bar{t}_j}^2$$

From this, Lemma 5.2, (5.71), and (5.72) it follows that

$$||(1-\psi)e_\sigma||_T^2 \leq \sum_{j \in \mathcal{J}} ||(1-\phi_j)e_q||_T^2, \quad \forall \ q \in \mathcal{P}$$

Thus for $q \in \mathcal{P}$

$$\begin{aligned} ||(1-\psi)e_\sigma||_T^2 &\leq \sum_{j \in \mathcal{J}} \int_0^T \{e_q e^{\lambda t}\}^2 (1-\phi_j)^2 dt \\ &= \int_0^T \{e_q e^{\lambda t}\}^2 \left\{ \sum_{j \in \mathcal{J}} (1-\phi_j)^2 \right\} dt \\ &= \int_0^T \{e_q e^{\lambda t}\}^2 \left\{ \sum_{j \in \mathcal{J}} (1-\phi_j) \right\} dt \end{aligned}$$

This, Lemma 5.3 and (5.70) imply that

$$||(1-\psi)e_\sigma||_T^2 \le \int_0^T \{e_q e^{\lambda t}\}^2 \left\{ m - 1 + \prod_{p\in\mathcal{P}_T} (1-\phi_p) \right\} dt$$
$$= \int_0^T \{e_q e^{\lambda t}\}^2 \{m - \psi\} dt$$
$$= \int_0^T \{e_q e^{\lambda t}\}^2 \{m - \psi^2\} dt$$

Hence

$$||(1-\psi)e_\sigma||_T^2 \le m||e_q||_T^2 - ||\psi e_q||_T^2 \tag{5.73}$$

Now

$$||(1-\psi)e_\sigma||_T^2 + ||\psi e_q||_T^2 = ||(1-\psi)e_\sigma + \psi e_q||_T^2$$

because $\psi(1-\psi) = 0$. From this and (5.73) it follows that

$$||(1-\psi)e_\sigma + \psi e_q||_T^2 \le m||e_q||_T^2$$

and thus that (5.28) is true. ■

Implication of Dwell - Time Switching When $\mathcal{P}$ is Not a Finite Set

To prove Proposition 5.2 we will need the following result which can be easily deduced from the discussion about strong bases in section 5.4.

Lemma 5.4. *Let ϵ be a positive number and suppose that $\mathcal{X} = \{x_1, x_2, \ldots x_m\}$ is any finite set of vectors in a real n-dimensional space such that $|x_m| > \epsilon$. There exists a subset of $\bar{m} \le n$ positive integers $\mathcal{N} = \{i_1, i_2, \ldots, i_{\bar{m}}\}$, each no larger than m, and a set of real numbers a_{ij}, $i \in \mathcal{M} = \{1, 2, \ldots m\}$, $j \in \mathcal{N}$ such that*

$$\left| x_i - \sum_{j\in\mathcal{N}} a_{ij} x_j \right| \le \epsilon, \quad i \in \mathcal{M}$$

where

$$a_{ij} = 0, \quad i \in \mathcal{M}, \quad j \in \mathcal{N}, \quad i > j$$
$$|a_{ij}| \le \frac{(1+\sup \mathcal{X})^n}{\epsilon}, \quad i \in \mathcal{M}, \quad j \in \mathcal{N}$$

Proof of Proposition 5.2: There are two cases to consider:
Case I: Suppose that $|c_{p_i q}| \le \rho$ for $i \in \{1, 2, \ldots, k\}$. In this case set $\psi(t) = 0$, $t \ge 0$, $h(t) = c_{\sigma(t)q}$ for $t \in [0, t_k)$, and $h(t) = 0$ for $t > t_k$. Then (5.47) and (5.46) hold and $e_\sigma = hx + e_q$ for $t \in [0, T)$. Therefore $||e_\sigma - hx||_T = ||e_q||_T$ and (5.48) follows.

Case II: Suppose the assumption of Case I is not true in which case there is a largest integer $m \in \{1, 2, \ldots, k\}$ such that $|c_{p_m q}| > \rho$. We claim that there is a non-negative integer $\bar{m} \leq n_E$, a set of $\bar{m}$ positive integers $\mathcal{J} = \{i_1, i_2, \ldots, i_{\bar{m}}\}$, each no greater than k, and a set of piecewise constant signals $\gamma_j : [0, \infty) \to \mathbb{R}$, $j \in \mathcal{J}$, such that

$$\left| c_{\sigma(t)q} - \sum_{j \in \mathcal{J}} \gamma_j(t) c_{p_j q} \right| \leq \rho, \quad 0 \leq t \leq T \tag{5.74}$$

where for all $j \in \mathcal{J}$

$$\gamma_j(t) = 0, \quad t \in (t_j, \infty) \tag{5.75}$$

$$|\gamma_j(t)| \leq \left(\frac{1 + \sup_{p \in \mathcal{P}} |c_{pq}|}{\rho} \right)^{n_E}, \quad t \in [0, t_j) \tag{5.76}$$

To establish this claim, we first note that $\{c_{p_1 q}, c_{p_2 q}, \ldots, c_{p_m q}\} \subset \{c_{pq} : p \in \mathcal{P}\}$ and that $\{c_{pq} : p \in \mathcal{P}\}$ is a bounded subset of an n_E dimensional space. By Lemma 5.4 we thus know that there must exist a subset of $\bar{m} \leq n_E$ integers $\mathcal{J} = \{i_1, i_2, \ldots, i_{\bar{m}}\}$, each no greater than m, and a set of real numbers g_{ij}, $i \in \mathcal{M} = \{1, 2, \ldots, m\}$, $j \in \mathcal{J}$ such that

$$\left| c_{p_i q} - \sum_{j \in \mathcal{J}} g_{ij} c_{p_j q} \right| \leq \rho, \quad i \in \mathcal{M} \tag{5.77}$$

where

$$g_{ij} = 0, \quad i \in \mathcal{M},\ j \in \mathcal{J},\ i > j, \tag{5.78}$$

$$|g_{ij}| \leq \frac{(1 + \sup_{p \in \mathcal{P}} |c_{pq}|)^{n_E}}{\rho}, \quad i \in \mathcal{M},\ j \in \mathcal{N} \tag{5.79}$$

Thus if for each $j \in \mathcal{J}$, we define $\gamma_j(t) = g_{ij}$, $t \in [t_{i-1}, t_i)$, $i \in \mathcal{M}$, and $\gamma_j(t) = 0$, $t > t_m$ then (5.74) - (5.76) will all hold.

To proceed, define $h(t)$ for $t \in [0, t_m)$ so that

$$h(t) = c_{p_i q} - \sum_{j \in \mathcal{J}} g_{ij} c_{p_j q},\ t \in [t_{i-1}, t_i),\ i \in \mathcal{M}$$

and for $t > t_m$ so that

$$h(t) = \begin{cases} c_{p_i q}\ t \in [t_{i-1}, t_i)\ i \in \{m+1, \ldots, k\} \\ \\ 0 \quad t > t_k \end{cases}$$

Then (5.46) holds because of (5.74) and the assumption that $|c_{p_i q}| \leq \rho$ for $i \in \{m+1, \ldots, k\}$. The definition of h implies that

$$c_{\sigma(t)q} - h(t) = \sum_{j\in\mathcal{J}} \gamma_j(t) c_{p_j q}, \quad t \in [0,T)$$

and thus that

$$e_{\sigma(t)}(t) - e_q(t) - h(t)x(t) = \sum_{j\in\mathcal{J}} \gamma_j(t)(e_{p_j}(t) - e_q(t)), \quad t \in [0,T) \tag{5.80}$$

For each $j \in \mathcal{J}$, let

$$\psi \triangleq 1 - \prod_{j\in\mathcal{J}} (1-\phi_j), \tag{5.81}$$

Since each ϕ_j has co-domain $\{0,1\}$, support no greater than $\tau_D + \tau_C$ and is idempotent, it must be true that ψ also has co-domain $\{0,1\}$, is idempotent, and has support no greater than $\bar{m}(\tau_D + \tau_C)$. In view of the latter property and the fact that $\bar{m} \le n_E$, (5.47) must be true.

By Lemma 5.2

$$||(1-\phi_j)e_{p_j}||_{\bar{t}_j} \le ||e_q||_T, \quad \forall j \in \mathcal{J}, \ \ q \in \mathcal{P}$$

From this and the triangle inequality

$$||(1-\phi_j)(e_{p_j} - e_q)||_{\bar{t}_j} \le 2||e_q||_T, \quad \forall j \in \mathcal{J}, \ \ q \in \mathcal{P} \tag{5.82}$$

From (5.80)

$$\begin{aligned} ||(1-\psi)(e_\sigma - e_q - \bar{f}x)||_T &= \left\| \textstyle\sum_{j\in\mathcal{N}} (1-\psi)\gamma_j(e_{p_j} - e_q) \right\|_T \\ &\le \textstyle\sum_{j\in\mathcal{N}} ||(1-\psi)\gamma_j(e_{p_j} - e_q)||_T \end{aligned} \tag{5.83}$$

But

$$||(1-\psi)\gamma_j(e_{p_j} - e_q)||_T = ||(1-\psi)\gamma_j(e_{p_j} - e_q)||_{t_j}$$

because of (5.75). In view of (5.76)

$$||(1-\psi)\gamma_j(e_{p_j} - e_q)||_{\{0,T\}} \le \bar{\gamma}||(1-\psi)(e_{p_j} - e_q)||_{t_j} \tag{5.84}$$

where

$$\bar{\gamma} \triangleq \frac{(1 + \sup_{p\in\mathcal{P}} |c_{pq}|)^{n_E}}{\rho}$$

Now $||(1-\psi)(e_{p_j} - e_q)||_{t_j} \le ||(1-\phi_j)(e_{p_j} - e_q)||_{t_j}$ because of (5.70). From this, (5.82) and (5.84) it follows that

$$||(1-\psi)\gamma_j(e_{p_j} - e_q)||_T \le 2\bar{\gamma}||e_q||_T$$

In view of (5.83) and the fact that $\bar{m} \le n_E$, it follows that

$$||(1-\psi)(e_\sigma - e_q - hx)||_T \le 2n_E\bar{\gamma}||e_q||_T$$

But

$$(1-\psi)(e_\sigma - hx)) + \psi e_q = (1-\psi)(e_\sigma - e_q - hx) - e_q$$

so by the triangle inequality

$$||(1-\psi)(e_\sigma - hx)) + \psi e_q||_T \leq ||(1-\psi)(e_\sigma - e_q - hx)||_T + ||e_q||_T$$

Therefore

$$||(1-\psi)(e_\sigma - hx) + \psi e_q||_T \leq (1 + 2n_E\bar{\gamma})||e_q||_T$$

and (5.48) is true. ■

Strong Bases

Let $\mathcal{X}$ be a subset of a real, finite dimensional linear space with norm $|\cdot|$ and let ϵ be a positive number. A nonempty list of vectors $\{x_1,\ x_2,\ \ldots,\ x_{\bar{n}}\}$ in $\mathcal{X}$ is ϵ-*independent* if

$$|x_{\bar{n}}| \geq \epsilon, \tag{5.85}$$

and, for $k \in \{1, 2, \ldots, \bar{n}-1\}$,

$$\left| x_k + \sum_{j=k+1}^{\bar{n}} \mu_j x_j \right| \geq \epsilon, \quad \forall \mu_j \in \mathbb{R} \tag{5.86}$$

$\{x_1,\ x_2,$
$ldots, x_{\bar{n}}\}$ ϵ-*spans* $\mathcal{X}$ if for each $x \in \mathcal{X}$ there is a set of real numbers $\{b_1, b_2, \ldots, b_{\bar{n}}\}$, called ϵ-*coordinates*, such that

$$\left| x - \sum_{i=1}^{\bar{n}} b_i x_i \right| \leq \epsilon \tag{5.87}$$

The following lemma gives an estimate on how large these ϵ-coordinates can be assuming $\mathcal{X}$ is abounded subset.

Lemma 5.5. *Let $\mathcal{X}$ be a bounded subset which is ϵ- spanned by an ϵ- independent list $\{x_1, x_2, \ldots, x_{\bar{n}}\}$. Suppose that x is a vector in $\mathcal{X}$ and that $b_1, b_2, \ldots b_{\bar{n}}$ is a set of ϵ- coordinates of x with respect to $\{x_1,\ x_2,\ \ldots,\ x_{\bar{n}}\}$. Then*

$$|b_i| \leq \left(1 + \frac{\sup \mathcal{X}}{\epsilon}\right)^{\bar{n}}, \quad i \in \{1, 2, \ldots, \bar{n}\} \tag{5.88}$$

This lemma will be proved in a moment.

Now suppose that $\mathcal{X}$ is an finite list of vectors $x_1,\ x_2,\ \ldots,\ x_m$ in a real n-dimensional vector space. Suppose, in addition, that $|x_m| \geq \epsilon$. It is possible to extract from $\mathcal{X}$ an ordered subset $\{x_{i_1}, x_{i_2}, \cdots, x_{i_{\bar{n}}}\}$, with $\bar{n} \leq n$, which is

ϵ-independent and which ϵ-spans $\mathcal{X}$. Moreover the i_j can always be chosen so that

$$i_1 < i_2 < i_3 < \cdots < i_{\bar{n}} = m \tag{5.89}$$

and also so that for suitably defined $b_{ij} \in \mathbb{R}$

$$\left| x_i - \sum_{j=k+1}^{\bar{n}} b_{ij} x_{i_j} \right| \leq \epsilon, i \in \{i_k + 1, i_k + 2, \ldots, i_{k+1}\}, k \in \{1, 2, \ldots, \bar{n} - 1\} \tag{5.90}$$

$$\left| x_i - \sum_{j=1}^{\bar{n}} b_{ij} x_{i_j} \right| \leq \epsilon, i \in \{1, 2, \ldots, k_1\} \tag{5.91}$$

In fact, the procedure for doing this is almost identical to the familiar procedure for extracting from $\{x_1,\ x_2,\ \ldots,\ x_m\}$, an ordered subset which is linearly independent {in the usual sense} and which spans the span of $\{x_1,\ x_2,\ \ldots,\ x_m\}$. The construction of interest here begins by defining an integer $j_1 \stackrel{\Delta}{=} m$. j_2 is then defined to be the greatest integer $j < j_1$ such that

$$|x_j - \mu x_{j_1}| \geq \epsilon \quad \forall \mu \in \mathbb{R},$$

if such an integer exists. If not, one defines $\bar{n} \stackrel{\Delta}{=} 1$ and $i_1 \stackrel{\Delta}{=} j_1$ and the construction is complete. If j_2 exists, j_3 is then defined to be the greatest integer $j < j_2$ such that

$$|x_j - \mu_1 x_{j_1} - \mu_2 x_{j_2}| \geq \epsilon \quad \forall \mu_i \in \mathbb{R},$$

if such an integer exists. If not, one defines $\bar{n} \stackrel{\Delta}{=} 2$ and $i_k \stackrel{\Delta}{=} j_{\bar{m}+1-k},\ k \in \{1, 2\}$ .... and so on. By this process one thus obtains an ϵ-independent, ϵ-spanning subset of $\mathcal{X}$ for which there exist numbers a_{ij} such that (5.89)-(5.91) hold. Since such $b_{ij},\ j \in \{1, 2, \ldots, \bar{n}\}$, are ϵ-coordinates of $x_i,\ i \in \{1, 2, \ldots, \bar{n}\}$, each coordinate must satisfy the same bound inequality as the b_i in (5.88). Moreover, because $\bar{n}$ cannot be larger than the dimension of the smallest linear space containing $\mathcal{X}$, $\bar{n} \leq n$. **Proof of Lemma 5.5:** For $k \in \{1, 2, \ldots, \bar{n}\}$ let

$$y_k \stackrel{\Delta}{=} \sum_{i=k}^{\bar{n}} b_i x_i \tag{5.92}$$

We claim that

$$|b_k| \leq \frac{|y_k|}{\epsilon}, \quad k \in \{1, 2, \ldots, \bar{n}\} \tag{5.93}$$

Now (5.93) surely holds for $k = \bar{n}$, because of (5.85) and the formula $|y_{\bar{n}}| = |b_{\bar{n}}||x_{\bar{n}}|$ which, in turn, is a consequence of (5.92). Next fix $k \in \{1, 2, , \ldots, \bar{n} - 1\}$. Now (5.93) is clearly true if $b_k = 0$. Suppose $b_k \neq 0$ in which case

$$y_k = b_k \left(x_k + \sum_{j=k+1}^{\bar{n}} \mu_j x_j \right)$$

where $\mu_j \stackrel{\Delta}{=} \frac{b_j}{b_k}$. From this and (5.86) it follows that $|y_k| \geq |b_k|\epsilon$, $k \in \{1, 2, \ldots, \bar{n}\}$, so (5.93) is true.

Next write $y_1 = (y_1 - x) + x$. Then $|y_1| \leq |y_1 - x| + |x|$. But $|x| \leq \sup \mathcal{X}$ because $x \in \mathcal{X}$ and $|y_1 - x| \leq \epsilon$ because of (5.87) and the definition of y_1 in (5.92). Therefore

$$\frac{|y_1|}{\epsilon} \leq \left(1 + \frac{\sup \mathcal{X}}{\epsilon} \right) \tag{5.94}$$

From (5.92) we have that $y_{k+1} = y_k - b_k x_k$, $k \in \{1, 2, \ldots, \bar{n} - 1\}$. Thus $|y_{k+1}| \leq |y_k| + |b_k||x_k|$, $k \in \{1, 2, \ldots, \bar{n} - 1\}$. Dividing both sides of this inequality by ϵ and then using (5.93) and $|x_k| \leq \sup \mathcal{X}$, we obtain the inequality

$$\frac{|y_{k+1}|}{\epsilon} \leq \left(1 + \frac{\sup \mathcal{X}}{\epsilon} \right) \frac{|y_k|}{\epsilon}, \quad k \in \{1, 2, \ldots, \bar{n} - 1\}$$

This and (5.94) imply that

$$\frac{|y_k|}{\epsilon} \leq \left(1 + \frac{\sup \mathcal{X}}{\epsilon} \right)^k, \quad k \in \{1, 2, \ldots, \bar{n}\}$$

In view of (5.93), it follows that (5.88) is true. ■

6 Flocking

Current interest in cooperative control of groups of mobile autonomous agents has led to the rapid increase in the application of graph theoretic ideas together with more familiar dynamical systems concepts to problems of analyzing and synthesizing a variety of desired group behaviors such as maintaining a formation, swarming, rendezvousing, or reaching a consensus. While this in-depth assault on group coordination using a combination of graph theory and system theory is in its early stages, it is likely to significantly expand in the years to come. One line of research which "graphically" illustrates the combined use of these concepts, is the recent theoretical work by a number of individuals which successfully explains the heading synchronization phenomenon observed in simulation by Vicsek [25], Reynolds [26] and others more than a decade ago. Vicsek and co-authors consider a simple discrete-time model consisting of n autonomous agents or particles all moving in the plane with the same speed but with different headings. Each agent's heading is updated using a local rule based on the average of its own heading plus the current headings of its "neighbors." Agent i's *neighbors* at time t, are those agents which are either in or on a circle of pre-specified radius r_i centered

at agent i's current position. In their paper, Vicsek *et al.* provide a variety of interesting simulation results which demonstrate that the nearest neighbor rule they are studying can cause all agents to eventually move in the same direction despite the absence of centralized coordination and despite the fact that each agent's set of nearest neighbors can change with time. A theoretical explanation for this observed behavior has recently been given in [27]. The explanation exploits ideas from graph theory [28] and from the theory of non-homogeneous Markov chains [29, 30, 31]. With the benefit of hind-sight it is now reasonably clear that it is more the graph theory than the Markov chains which will prove key as this line of research advances. An illustration of this is the recent extension of the findings of [27] which explain the behavior of Reynolds' full nonlinear "boid" system [32]. By appealing to the concept of *graph composition*, we side-step most issues involving products of stochastic matrices and present in this chapter a variety of graph theoretic results which explain how convergence to a common heading is achieved.

Since the writing of [27] many important papers have appeared which extend the Vicsek problem in many directions and expand the results obtained [33, 34, 35, 36, 37]. Especially noteworthy among these are recent papers by Moreau [33] and Beard [34] which address the modified versions of the Vicsek problem in which different agents use different sensing radii r_i. The asymmetric neighbor relationships which result necessitate the use of directed graphs rather than undirected graphs to represent neighbor relation. We will use directed graphs in this chapter, not only because we want to deal with different sensing radii, but also because working with directed graphs enables us to give convergence conditions for the symmetric version of Vicsek's problem which are less restrictive than those originally presented in [27].

Vicsek's problem is what in computer science is called a "consensus problem" or an "agreement problem." Roughly speaking, one has a group of agents which are all trying to agree on a specific value of some quantity. Each agent initially has only limited information available. The agents then try to reach a consensus by passing what they know between them either just once or repeatedly, depending on the specific problem of interest. For the Vicsek problem, each agent always knows only its own heading and the headings of its neighbors. One feature of the Vicsek problem which sharply distinguishes it from other consensus problems, is that each agent's neighbors change with time, because all agents are in motion for the problems considered in these notes. The theoretical consequence of this is profound: it renders essentially useless a large body of literature appropriate to the convergence analysis of "nearest neighbor" algorithms with fixed neighbor relationships. Said differently, for the linear heading update rules considered in this chapter, understanding the difference between fixed neighbor relationships and changing neighbor relationships is much the same as understanding the difference between the stability of time - invariant linear systems and time - varying linear systems.

6.1 Leaderless Coordination

The system to be studied consists of n autonomous agents, labelled 1 through n, all moving in the plane with the same speed but with different headings. Each agent's heading is updated using a simple local rule based on the average of its own heading plus the headings of its "neighbors." Agent i's *neighbors* at time t, are those agents, including itself, which are either in or on a circle of pre-specified radius r_i centered at agent i's current position. In the sequel $\mathcal{N}_i(t)$ denotes the set of labels of those agents which are neighbors of agent i at time t. Agent i's heading, written θ_i, evolves in discrete-time in accordance with a model of the form

$$\theta_i(t+1) = \frac{1}{n_i(t)} \left(\sum_{j \in \mathcal{N}_i(t)} \theta_j(t) \right) \tag{6.1}$$

where t is a discrete-time index taking values in the non-negative integers $\{0, 1, 2, \ldots\}$, and $n_i(t)$ is the number of neighbors of agent i at time t.

The explicit form of the update equations determined by (6.1) depends on the relationships between neighbors which exist at time t. These relationships can be conveniently described by a directed graph $\mathbb{G}$ with vertex set $\mathcal{V} = \{1, 2, \ldots n\}$ and "arc set" $\mathcal{A}(\mathbb{G}) \subset \mathcal{V} \times \mathcal{V}$ which is defined in such a way so that (i, j) is an *arc* or directed edge from i to j just in case agent i is a neighbor of agent j. Thus $\mathbb{G}$ is a directed graph on n vertices with at most one arc from any vertex to another and with exactly one self - arc at each vertex. We write $\mathcal{G}$ for the set of all such graphs and $\bar{\mathcal{G}}$ for the set of all directed graphs with vertex set $\mathcal{V}$. We use the symbol $\bar{\mathcal{P}}$ to denote a suitably defined set indexing $\bar{\mathcal{G}}$ and we write $\mathcal{P}$ for the subset of $\bar{\mathcal{P}}$ which indexes $\mathcal{G}$. Thus $\mathcal{G} = \{\mathbb{G}_p : p \in \mathcal{P}\}$ where for $p \in \bar{\mathcal{P}}$, $\mathbb{G}_p$ denotes the pth graph in $\bar{\mathcal{G}}$. It is natural to call to a vertex i a *neighbor* of vertex j in $\mathbb{G}$ if (i, j) is and arc in $\mathbb{G}$. In addition we sometimes refer to a vertex k as a *observer* of vertex j in $\mathbb{G}$ if (j, k) is and arc in $\mathbb{G}$. Thus every vertex of $\mathbb{G}$ can *observe* its neighbors, which with the interpretation of vertices as agents, is precisely the kind of relationship $\mathbb{G}$ is suppose to represent.

The set of agent heading update rules defined by (6.1) can be written in state form. Toward this end, for each $p \in \mathcal{P}$, define *flocking matrix*

$$F_p = D_p^{-1} A_p' \tag{6.2}$$

where A_p' is the transpose of the "adjacency matrix" of the graph $\mathbb{G}_p$ and D_p the diagonal matrix whose jth diagonal element is the "in-degree" of vertex j within the graph[4]. Then

[4] By the *adjacency matrix* of a directed graph $\mathbb{G} \in \bar{\mathcal{G}}$ is meant an $n \times n$ matrix whose ijth entry is a 1 if (i, j) is an arc in $\mathcal{A}(\mathbb{G})$ and 0 if it is not. The *in-degree* of vertex j in $\mathbb{G}$ is the number of arcs in $\mathcal{A}(\mathbb{G})$ of the form (i, j); thus j's in-degree is the number of *incoming* arcs to vertex j.

$$\theta(t+1) = F_{\sigma(t)}\theta(t), \quad t \in \{0,1,2,\ldots\} \tag{6.3}$$

where θ is the heading vector $\theta = \begin{pmatrix}\theta_1 & \theta_2 & \ldots & \theta_n\end{pmatrix}'$ and $\sigma : \{0,1,\ldots\} \to \mathcal{P}$ is a switching signal whose value at time t, is the index of the graph representing the agents' neighbor relationships at time t. A complete description of this system would have to include a model which explains how σ changes over time as a function of the positions of the n agents in the plane. While such a model is easy to derive and is essential for simulation purposes, it would be difficult to take into account in a convergence analysis. To avoid this difficulty, we shall adopt a more conservative approach which ignores how σ depends on the agent positions in the plane and assumes instead that σ might be any switching signal in some suitably defined set of interest.

Our ultimate goal is to show for a large class of switching signals and for any initial set of agent headings that the headings of all n agents will converge to the same steady state value θ_{ss}. Convergence of the θ_i to θ_{ss} is equivalent to the state vector θ converging to a vector of the form $\theta_{ss}\mathbf{1}$ where $\mathbf{1} \stackrel{\Delta}{=} \begin{pmatrix}1 & 1 & \ldots & 1\end{pmatrix}'_{n\times 1}$. Naturally there are situations where convergence to a common heading cannot occur. The most obvious of these is when one agent - say the ith - starts so far away from the rest that it never acquires any neighbors. Mathematically this would mean not only that $\mathbb{G}_{\sigma(t)}$ is never strongly connected[5] at any time t, but also that vertex i remains an isolated vertex of $\mathbb{G}_{\sigma(t)}$ for all t in the sense that within each $\mathbb{G}_{\sigma(t)}$, vertex i has no incoming arcs other than its own self - arc. This situation is likely to be encountered if the r_i are very small. At the other extreme, which is likely if the r_i are very large, all agents might remain neighbors of all others for all time. In this case, σ would remain fixed along such a trajectory at that value in $p \in \mathcal{P}$ for which $\mathbb{G}_p$ is a complete graph. Convergence of θ to $\theta_{ss}\mathbf{1}$ can easily be established in this special case because with σ so fixed, (6.3) is a linear, time-invariant, discrete-time system. The situation of perhaps the greatest interest is between these two extremes when $\mathbb{G}_{\sigma(t)}$ is not necessarily complete or even strongly connected for any $t \geq 0$, but when no strictly proper subset of $\mathbb{G}_{\sigma(t)}$'s vertices is isolated from the rest for all time. Establishing convergence in this case is challenging because σ changes with time and (6.3) is not time-invariant. It is this case which we intend to study.

Strongly Rooted Graphs

In the sequel we will call a vertex i of a directed graph $\mathbb{G} \in \bar{\mathcal{G}}$, a *root* of $\mathbb{G}$ if for each other vertex j of $\mathbb{G}$, there is a path from i to j. Thus i is a root of $\mathbb{G}$,

[5] A directed graph $\mathbb{G} \in \bar{\mathcal{G}}$ with arc set $\mathcal{A}$ is *strongly connected* if has a "path" between each distinct pair of its vertices i and j; by a *path* {of *length* m} between vertices i and j is meant a sequence of arcs in $\mathcal{A}$ of the form $(i,k_1),(k_1,k_2),\ldots(k_m,j)$ where $i,k_1,\ldots,k_m$, and j are distinct vertices. $\mathbb{G}$ is *complete* if has a path of length one {i.e., an arc} between each distinct pair of its vertices.

if it is the root of a directed spanning tree of $\mathbb{G}$. We will say that $\mathbb{G}$ is *rooted at i* if i is in fact a root. Thus $\mathbb{G}$ is rooted at i just in case each other vertex of $\mathbb{G}$ is *reachable* from vertex i along a path within the graph. $\mathbb{G}$ is *strongly rooted at i* if each other vertex of $\mathbb{G}$ is reachable from vertex i along a path of length 1. Thus $\mathbb{G}$ is strongly rooted at i if i is a neighbor of every other vertex in the graph. By a *rooted graph* $\mathbb{G} \in \bar{\mathcal{G}}$ is meant a graph which possesses at least one root. Finally, a *strongly rooted graph* is a graph which at least one vertex at which it is strongly rooted. It is now possible to state the following elementary convergence result.

Theorem 6.1. *Let $\mathcal{Q}$ denote the subset of $\mathcal{P}$ consisting of those indices q for which $\mathbb{G}_q \in \mathcal{G}$ is strongly rooted. Let $\theta(0)$ be fixed and let $\sigma : \{0, 1, 2, \ldots\} \to \mathcal{P}$ be a switching signal satisfying $\sigma(t) \in \mathcal{Q}$, $t \in \{0, 1, \ldots\}$. Then there is a constant steady state heading θ_{ss} depending only on $\theta(0)$ and σ for which*

$$\lim_{t \to \infty} \theta(t) = \theta_{ss}\mathbf{1} \tag{6.4}$$

where the limit is approached exponentially fast.

In order to explain why this theorem is true, we will make use of certain structural properties of the F_p. As defined, each F_p is square and non-negative, where by a *non-negative* matrix is meant a matrix whose entries are all non-negative. Each F_p also has the property that its row sums all equal 1 {i.e., $F_p\mathbf{1} = \mathbf{1}$}. Matrices with these two properties are called {row} *stochastic* [38]. Because each vertex of each graph in $\mathcal{G}$ has a self-arc, the F_p have the additional property that their diagonal elements are all non-zero. Let $\mathcal{S}$ denote the set of all $n \times n$ row stochastic matrices whose diagonal elements are all positive. $\mathcal{S}$ is closed under multiplication because the class of all $n \times n$ stochastic matrices is closed under multiplication and because the class of $n \times n$ non-negative matrices with positive diagonals is also.

In the sequel we write $M \geq N$ whenever $M - N$ is a non-negative matrix. We also write $M > N$ whenever $M - N$ is a positive matrix where by a *positive matrix* is meant a matrix with all positive entries.

Products of Stochastic Matrices

Stochastic matrices have been extensively studied in the literature for a long time largely because of their connection with Markov chains [29, 30, 31]. One problem studied which is of particular relevance here, is to describe the asymptotic behavior of products of $n \times n$ stochastic matrices of the form

$$S_j S_{j-1} \cdots S_1$$

as j tends to infinity. This is equivalent to looking at the asymptotic behavior of all solutions to the recursion equation

$$x(j+1) = S_j x(j) \tag{6.5}$$

since any solution $x(j)$ can be written as

$$x(j) = (S_j S_{j-1} \cdots S_1)\, x(1), \quad j \geq 1$$

One especially useful idea, which goes back at least to [39] and more recently to [40], is to consider the behavior of the scalar-valued non-negative function $V(x) = \lceil x \rceil - \lfloor x \rfloor$ along solutions to (6.5) where $x = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}'$ is a non-negative n vector and $\lceil x \rceil$ and $\lfloor x \rfloor$ are its largest and smallest elements respectively. The key observation is that for any $n \times n$ stochastic matrix S, the ith entry of Sx satisfies

$$\sum_{j=1}^{n} s_{ij} x_j \geq \sum_{j=1}^{n} s_{ij} \lfloor x \rfloor = \lfloor x \rfloor$$

and

$$\sum_{j=1}^{n} s_{ij} x_j \leq \sum_{j=1}^{n} s_{ij} \lceil x \rceil = \lceil x \rceil$$

Since these inequalities hold for all rows of Sx, it must be true that $\lfloor Sx \rfloor \geq \lfloor x \rfloor$, $\lceil Sx \rceil \leq \lceil x \rceil$ and, as a consequence, that $V(Sx) \leq V(x)$. These inequalities and (6.5) imply that the sequences

$$\lfloor x(1) \rfloor, \lfloor x(2) \rfloor, \ldots \qquad \lceil x(1) \rceil, \lceil x(2) \rceil, \ldots \qquad V(x(1)), V(x(2)), \ldots$$

are each monotone. Thus because each of these sequences is also bounded, the limits

$$\lim_{j\to\infty} \lfloor x(j) \rfloor, \qquad \lim_{j\to\infty} \lceil x(j) \rceil, \qquad \lim_{j\to\infty} V(x(j))$$

each exist. Note that whenever the limit of $V(x(j))$ is zero, all components of $x(j)$ must tend to the same value and moreover this value must be a constant equal to the limiting value of $\lfloor x(j) \rfloor$.

There are various different ways one might approach the problem of developing conditions under which $S_j S_{j-1} \cdots S_1$ converges to a constant matrix of the form $\mathbf{1}c$ or equivalently $x(j)$ converges to some scalar multiple of $\mathbf{1}$. For example, since for any $n \times n$ stochastic matrix S, $S\mathbf{1} = \mathbf{1}$, it must be true that span $\{\mathbf{1}\}$ is an S-invariant subspace for any such S. From this and standard existence conditions for solutions to linear algebraic equations, it follows that for any $(n-1) \times n$ matrix P with kernel spanned by $\mathbf{1}$, the equations $PS = \tilde{S}P$ has unique solutions $\tilde{S}$, and moreover that

$$\text{spectrum } S = \{1\} \cup \text{spectrum } \tilde{S} \tag{6.6}$$

As a consequence of the equations $PS_j = \tilde{S}_j P, \;\; j \geq 1$, it can easily be seen that

$$\tilde{S}_j \tilde{S}_{j-1} \cdots \tilde{S}_1 P = P S_j S_{j-1} \cdots S_1$$

Since P has full row rank and $P\mathbf{1} = 0$, the convergence of a product of the form $S_j S_{j-1} \cdots S_1$ to $\mathbf{1}c$ for some constant row vector c, is equivalent to convergence

of the corresponding product $\tilde{S}_j\tilde{S}_{j-1}\cdots\tilde{S}_1$ to the zero matrix. There are two problems with this approach. First, since P is not unique, neither are the $\tilde{S}_i$. Second it is not so clear how to going about picking P to make tractable the problem of proving that the resulting product $\tilde{S}_j\tilde{S}_{j-1}\cdots\tilde{S}_1$ tends to zero. Tractability of the latter problem generally boils down to choosing a norm for which the $\tilde{S}_i$ are all contractive. For example, one might seek to choose a suitably weighted 2-norm. This is in essence the same thing choosing a common quadratic Lyapunov function. Although each $\tilde{S}_i$ can easily be shown to be discrete-time stable, it is known that there are classes of S_i which give rise to $\tilde{S}_i$ for which no such common Lyapunov matrix exists [41] regardless of the choice of P. Of course there are many other possible norms to choose from other than 2 norms. In the end, success with this approach requires one to simultaneously choose *both* a suitable P and an appropriate norm with respect to which the $\tilde{S}_i$ are all contractive. In the sequel we adopt a slightly different, but closely related approach which ensures that we can work with what is perhaps the most natural norm for this type of convergence problem, the infinity norm.

To proceed, we need a few more ideas concerned with non-negative matrices. For any non-negative matrix R of any size, we write $||R||$ for the largest of the row sums of R. Note that $||R||$ is the induced infinity norm of R and consequently is sub-multiplicative. Note in addition that $||x|| = \lceil x \rceil$ for any non-negative n vector x. Moreover, $||M_1|| \leq ||M_2||$ if $M_1 \leq M_2$. Observe that for any $n \times n$ stochastic matrix S, $||S|| = 1$ because the row sums of a stochastic matrix all equal 1. We extend the domain of definitions of $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ to the class of all non-negative $n \times m$ matrix M, by letting $\lfloor M \rfloor$ and $\lceil M \rceil$ now denote the $1 \times m$ row vectors whose jth entries are the smallest and largest elements respectively, of the jth column of M. Note that $\lfloor M \rfloor$ is the largest $1 \times m$ non-negative row vector c for which $M - \mathbf{1}c$ is non-negative and that $\lceil M \rceil$ is the smallest non-negative row vector c for which $\mathbf{1}c - M$ is non-negative. Note in addition that for any $n \times n$ stochastic matrix S,

$$S = \mathbf{1}\lfloor S \rfloor + \lfloor\!\lfloor S \rfloor\!\rfloor \qquad \text{and} \qquad S = \mathbf{1}\lceil S \rceil - \lceil\!\lceil S \rceil\!\rceil \tag{6.7}$$

where $\lfloor\!\lfloor S \rfloor\!\rfloor$ and $\lceil\!\lceil S \rceil\!\rceil$are the non-negative matrices

$$\lfloor\!\lfloor S \rfloor\!\rfloor = S - \mathbf{1}\lfloor S \rfloor \qquad \text{and} \qquad \lceil\!\lceil S \rceil\!\rceil = \mathbf{1}\lceil S \rceil - S \tag{6.8}$$

respectively. Moreover the row sums of $\lfloor\!\lfloor S \rfloor\!\rfloor$ are all equal to $1 - \lfloor S \rfloor\mathbf{1}$ and the row sums of $\lceil\!\lceil S \rceil\!\rceil$ are all equal to $\lceil S \rceil\mathbf{1} - 1$ so

$$||\lfloor\!\lfloor S \rfloor\!\rfloor|| = 1 - \lfloor S \rfloor\mathbf{1} \qquad \text{and} \qquad ||\lceil\!\lceil S \rceil\!\rceil|| = \lceil S \rceil\mathbf{1} - 1 \tag{6.9}$$

In the sequel we will also be interested in the matrix

$$\lfloor\!\lceil S \rceil\!\rfloor = \lfloor\!\lfloor S \rfloor\!\rfloor + \lceil\!\lceil S \rceil\!\rceil \tag{6.10}$$

This matrix satisfies

$$[\![S]\!] = \mathbf{1}(\lceil S \rceil - \lfloor S \rfloor) \tag{6.11}$$

because of (6.7).

For any infinite sequence of $n \times n$ stochastic matrices $S_1, S_2, \ldots$, we henceforth use the symbol $\lfloor \cdots S_j \cdots S_1 \rfloor$ to denote the limit

$$\lfloor \cdots S_j \cdots S_2 S_1 \rfloor = \lim_{j \to \infty} \lfloor S_j \cdots S_2 S_1 \rfloor \tag{6.12}$$

From the preceding discussion it is clear that this limit exists whether or not the product $S_j \cdots S_2 S_1$ itself has a limit. Two situations can occur. Either the product $S_j \cdots S_2 S_1$ converges to a rank one matrix or it does not. It is quite possible for such a product to converge to a matrix which is not rank one. An example of this would be a sequence in which S_1 is any stochastic matric of rank greater than 1 and for all $i > 1$, $S_i = I_{n \times n}$. In the sequel we will develop necessary conditions for $S_j \cdots S_2 S_1$ to converge to a rank one matrix as $j \to \infty$. Note that if this occurs, then the limit must be of the form $\mathbf{1}c$ where $c\mathbf{1} = 1$ because stochastic matrices are closed under multiplication.

In the sequel we will say that a matrix product $S_j S_{j-1} \cdots S_1$ *converges to* $\mathbf{1}\lfloor \cdots S_j \cdots S_1 \rfloor$ *exponentially fast at a rate no slower than* λ if there are non-negative constants b and λ with $\lambda < 1$, such that

$$||(S_j \cdots S_1) - \mathbf{1}\lfloor \cdots S_j \cdots S_2 S_1 \rfloor|| \leq b\lambda^j, \qquad j \geq 1 \tag{6.13}$$

The following proposition implies that such a stochastic matrix product will so converge if the matrix product $[\![S_j \cdots S_1]\!]$ converges to 0.

Proposition 6.1. *Let $\bar{b}$ and λ be non-negative numbers with $\lambda < 1$. Suppose that $S_1, S_2, \ldots,$ is an infinite sequence of $n \times n$ stochastic matrices for which*

$$||[\![S_j \cdots S_1]\!]|| \leq \bar{b}\lambda^j, \quad j \geq 0 \tag{6.14}$$

Then the matrix product $S_j \cdots S_2 S_1$ converges to $\mathbf{1}\lfloor \cdots S_j \cdots S_1 \rfloor$ exponentially fast at a rate no slower than λ.

The proof of Proposition 6.1 makes use of the first of the two inequalities which follow.

Lemma 6.1. *For any two $n \times n$ stochastic matrices S_1 and S_2,*

$$\lfloor S_2 S_1 \rfloor - \lfloor S_1 \rfloor \leq \lceil S_2 \rceil [\![S_1]\!] \tag{6.15}$$

$$[\![S_2 S_1]\!] \leq [\![S_2]\!]\ [\![S_1]\!] \tag{6.16}$$

Proof of Lemma 6.1: Since $S_2 S_1 = S_2(\mathbf{1}\lfloor S_1 \rfloor + [\![S_1]\!]) = \mathbf{1}\lfloor S_1 \rfloor + S_2[\![S_1]\!]$ and $S_2 = \mathbf{1}\lceil S_2 \rceil - [\![S_2]\!]$, it must be true that $S_2 S_1 = \mathbf{1}(\lfloor S_1 \rfloor + \lceil S_2 \rceil [\![S_1]\!]) - [\![S_2]\!]\ [\![S_1]\!]$. But $\lceil S_2 S_1 \rceil$ is the smallest non-negative row vector c for which $\mathbf{1}c - S_2 S_1$ is non-negative. Therefore

$$\lceil S_2 S_1 \rceil \leq \lfloor S_1 \rfloor + \lceil S_2 \rceil \lfloor S_1 \rfloor \tag{6.17}$$

Moreover $\lfloor S_2 S_1 \rfloor \leq \lceil S_2 S_1 \rceil$ because of (6.11). This and (6.17) imply $\lfloor S_2 S_1 \rfloor \leq \lfloor S_1 \rfloor + \lceil S_2 \rceil \lfloor S_1 \rfloor$ and thus that (6.15) is true.

Since $S_2 S_1 = S_2(\mathbf{1}\lfloor S_1 \rfloor + \lfloor S_1 \rfloor) = \mathbf{1}\lfloor S_1 \rfloor + S_2 \lfloor S_1 \rfloor$ and $S_2 = \lfloor S_2 \rfloor + \lfloor S_2 \rfloor$, it must be true that $S_2 S_1 = \mathbf{1}(\lfloor S_1 \rfloor + \lfloor S_2 \rfloor \lfloor S_1 \rfloor) + \lfloor S_2 \rfloor \ \lfloor S_1 \rfloor$. But $\lfloor S_2 S_1 \rfloor$ is the largest non-negative row vector c for which $S_2 S_1 - \mathbf{1}c$ is non-negative so

$$S_2 S_1 \leq \mathbf{1}\lfloor S_2 S_1 \rfloor + \lfloor S_2 \rfloor \ \lfloor S_1 \rfloor \tag{6.18}$$

Now it is also true that $S_2 S_1 = \mathbf{1}\lfloor S_2 S_1 \rfloor + \lfloor S_2 S_1 \rfloor$. From this and (6.18) it follows that (6.16) is true. ■

Proof of Proposition 6.1: Set $X_j = S_j \cdots S_1, \quad j \geq 1$ and note that each X_j is a stochastic matrix. In view of (6.15),

$$\lfloor X_{j+1} \rfloor - \lfloor X_j \rfloor \leq \lceil S_{j+1} \rceil \lfloor X_j \rfloor, \quad j \geq 1$$

By hypothesis, $|| \lfloor X_j \rfloor || \leq \bar{b}\lambda^j, \quad j \geq 1$. Moreover $|| \lceil S_{j+1} \rceil || \leq n$ because all entries in S_{j+1} are bounded above by 1. Therefore

$$|| \lfloor X_{j+1} \rfloor - \lfloor X_j \rfloor || \leq n\bar{b}\lambda^j, \quad j \geq 1 \tag{6.19}$$

Clearly

$$\lfloor X_{j+i} \rfloor - \lfloor X_j \rfloor = \sum_{k=1}^{i} (\lfloor X_{i+j+1-k} \rfloor - \lfloor X_{i+j-k} \rfloor), \ \ i, j \geq 1$$

Thus, by the triangle inequality

$$|| \lfloor X_{j+i} \rfloor - \lfloor X_j \rfloor || \leq \sum_{k=1}^{i} || \lfloor X_{i+j+1-k} \rfloor - \lfloor X_{i+j-k} \rfloor ||, \ \ i, j \geq 1$$

This and (6.19) imply that

$$|| \lfloor X_{j+i} \rfloor - \lfloor X_j \rfloor || \leq n\bar{b} \sum_{k=1}^{i} \lambda^{(i+j-k)}, \quad i, j \geq 1$$

Now

$$\sum_{k=1}^{i} \lambda^{(i+j-k)} = \lambda^j \sum_{k=1}^{i} \lambda^{(i-k)} = \lambda^j \sum_{q=1}^{i} \lambda^{q-1} \leq \lambda^j \sum_{q=1}^{\infty} \lambda^{q-1}$$

But $\lambda < 1$ so

$$\sum_{q=1}^{\infty} \lambda^{q-1} = \frac{1}{(1-\lambda)}$$

Therefore

$$||\lfloor X_{i+j}\rfloor - \lfloor X_j\rfloor|| \leq n\bar{b}\frac{\lambda^j}{(1-\lambda)}, \quad i,j \geq 1 \tag{6.20}$$

Set $c = \lfloor \cdots S_j \cdots S_1 \rfloor$ and note that

$$\begin{aligned} ||\lfloor X_j\rfloor - c|| &= ||\lfloor X_j\rfloor - \lfloor X_{i+j}\rfloor + \lfloor X_{i+j}\rfloor - c|| \\ &\leq ||\lfloor X_j\rfloor - \lfloor X_{i+j}\rfloor|| + ||\lfloor X_{i+j}\rfloor - c||, \quad i,j \geq 1 \end{aligned}$$

In view of (6.20)

$$||\lfloor X_j\rfloor - c|| \leq n\bar{b}\frac{\lambda^j}{(1-\lambda)} + ||\lfloor X_{i+j}\rfloor - c||, \quad i,j \geq 1$$

Since

$$\lim_{i\to\infty} ||\lfloor X_{i+j}\rfloor - c|| = 0$$

it must be true that

$$||\lfloor X_j\rfloor - c|| \leq n\bar{b}\frac{\lambda^j}{(1-\lambda)}, \quad j \geq 1$$

But $||\mathbf{1}(\lfloor X_j\rfloor - c)|| = ||\lfloor X_j\rfloor - c||$ and $X_j = S_j \cdots S_1$. Therefore

$$||\mathbf{1}(\lfloor S_j \cdots S_1\rfloor - c)|| \leq n\bar{b}\frac{\lambda^j}{(1-\lambda)}, \quad j \geq 1 \tag{6.21}$$

In view of (6.7)

$$S_j \cdots S_1 = \mathbf{1}\lfloor S_j \cdots S_1\rfloor + \lbag S_j \cdots S_1 \rbag, \quad j \geq 1$$

Therefore

$$\begin{aligned} ||(S_j \cdots S_1) - \mathbf{1}c|| &= ||\mathbf{1}\lfloor S_j \cdots S_1\rfloor + \lbag S_j \cdots S_1 \rbag - \mathbf{1}c|| \\ &\leq ||\mathbf{1}\lfloor S_j \cdots S_1\rfloor - \mathbf{1}c|| + ||\lbag S_j \cdots S_1 \rbag||, \quad j \geq 1 \end{aligned}$$

From this, (6.14) and (6.21) it follows that

$$||S_j \cdots S_1 - \mathbf{1}c|| \leq \bar{b}\left(1 + \frac{n}{(1-\lambda)}\right)\lambda^j, \quad j \geq 1$$

and thus that (6.13) holds with $b = \bar{b}\left(1 + \frac{n}{(1-\lambda)}\right)$. ■

Convergence

We are now in a position to make some statements about the asymptotic behavior of a product of $n \times n$ stochastic matrices of the form $S_j S_{j-1} \cdots S_1$ as

j tends to infinity. Note first that (6.16) generalizes to sequences of stochastic matrices of any length. Thus

$$\lfloor S_j S_{j-1} \cdots S_2 S_1 \rfloor \leq \lfloor S_j \rfloor \ \lfloor S_{j-1} \rfloor \cdots \lfloor S_1 \rfloor \tag{6.22}$$

It is therefore clear that condition (6.14) of Proposition 6.1 will hold with $\bar{b} = 1$ if

$$||\lfloor S_j \rfloor \cdots \lfloor S_1 \rfloor|| \leq \lambda^j \tag{6.23}$$

for some nonnegative number $\lambda < 1$. Because $||\cdot||$ is sub-multiplicative, this means that a product of stochastic matrices $S_j \cdots S_1$ will converge to a limit of the form $\mathbf{1}c$ for some constant row-vector c if for each of the matrices S_i in the sequence $S_1, S_2, \ldots$ satisfies the norm bound $||\lfloor S_i \rfloor|| < \lambda$. We now develop a condition, tailored to our application, for this to be so. For any $n \times n$ stochastic matrix, let $\gamma(S)$ denote that graph $\mathbb{G} \in \bar{\mathcal{G}}$ whose adjacency matrix is the transpose of the matrix obtained by replacing all of S's non-zero entries with 1s.

Lemma 6.2. *For each $n \times n$ stochastic matrix S whose graph $\gamma(S)$ is strongly rooted*

$$||\lfloor S \rfloor|| < 1 \tag{6.24}$$

Proof: Let A be the adjacency matrix of $\gamma(S)$. Since $\gamma(S)$ is strongly rooted, its adjacency matrix A must have a positive row i for every i which is the label of a root of $\gamma(S)$. Since the positions of the non-zero entries of S and A' are the same, this means that S's ith column s_i will be positive if i is a root. Clearly $\lfloor S \rfloor$ will have its ith entry non-zero if vertex i is a root of $\gamma(S)$. Since $\gamma(S)$ is strongly rooted, at least one such root exists which implies that $\lfloor S \rfloor$ is non-zero, and thus that $1 - \lfloor S \rfloor \mathbf{1} < 1$. From this and (6.9) it follows that (6.24) is true. ■

It can be shown very easily that if (6.24) holds, then $\gamma(S)$ must be strongly rooted. We will not need this fact so the proof is omitted.

Proposition 6.2. *Let $\mathcal{S}_{sr}$ be any closed set of $n \times n$ stochastic matrices whose graphs $\gamma(S)$, $S \in \mathcal{S}_{sr}$ are all strongly rooted. Then any product $S_j \cdots S_1$ of matrices from $\mathcal{S}_{sr}$ converges to $\mathbf{1}\lfloor \cdots S_j \cdots S_1 \rfloor$ exponentially fast as $j \to \infty$ at a rate no slower than λ, where λ is a non-negative constant depending on $\mathcal{S}_{sr}$ and satisfying $\lambda < 1$.*

Proof of Proposition 6.2: In view of Lemma 6.2, $||\lfloor S \rfloor|| < 1$, $S \in \mathcal{S}_{sr}$. Let

$$\lambda = \max_{S \in \mathcal{S}_{sr}} ||\lfloor S \rfloor||$$

Because $\mathcal{S}_{sr}$ is closed and bounded and $||\lfloor \cdot \rfloor||$ is continuous, $\lambda < 1$. Clearly $||\lfloor S \rfloor|| \leq \lambda$, $i \geq 1$ so (6.23) must hold for any sequence of matrices $S_1, S_2, \ldots$ from $\mathcal{S}_{sr}$. Therefore for any such sequence $||\lfloor S_j \cdots S_1 \rfloor|| \leq \lambda^j$, $j \geq 0$. Thus by

Proposition 6.1, the product $\Pi(j) = S_j S_{j-1} \cdots S_1$ converges to $\mathbf{1}\lfloor \cdots S_j \cdot S_1 \rfloor$ exponentially fast at a rate no slower than λ. ■

Proof of Theorem 6.1: By definition, the graph $\mathbb{G}_p$ of each matrix F_p in the finite set $\{F_p : p \in \mathcal{Q}\}$ is strongly rooted. By assumption, $F_{\sigma(t)} \in \{F_p : p \in \mathcal{Q}\}$, $t \geq 0$. In view of Proposition 6.2, the product $F_{\sigma(t)} \cdots F_{\sigma(0)}$ converges to $\mathbf{1}\lfloor \cdots F_{\sigma(t)} \cdots F_{\sigma(0)} \rfloor$ exponentially fast at a rate no slower than

$$\lambda = \max_{p \in \mathcal{Q}} ||\lfloor F_p \rfloor||$$

But it is clear from (6.3) that

$$\theta(t) = F_{\sigma(t-1)} \cdots F_{\sigma(1)} F_{\sigma(0)} \theta(0), \quad t \geq 1$$

Therefore (6.4) holds with $\theta_{ss} = \lfloor \cdots F_{\sigma(t)} \cdots F_{\sigma(0)} \rfloor \theta(0)$ and the convergence is exponential. ■

Convergence Rate

Using (6.9) it is possible to calculate a worst case value for the convergence rate λ used in the proof of Theorem 6.1. Fix $p \in \mathcal{Q}$ and consider the flocking matrix F_p and its associated graph $\mathbb{G}_p$. Because $\mathbb{G}_p$ is strongly rooted, at least one vertex – say the kth – must be a root with arcs to each other vertex. In the context of (6.1), this means that agent k must be a neighbor of every agent. Thus θ_k must be in each sum in (6.1). Since each n_i in (6.1) is bounded above by n, this means that the smallest element in column k of F_p, is bounded below by $\frac{1}{n}$. Since (6.9) asserts that $||\lfloor F_p \rfloor|| = 1 - \lfloor F_p \rfloor \mathbf{1}$, it must be true that $||\lfloor F_p \rfloor|| \leq 1 - \frac{1}{n}$. This holds for all $p \in \mathcal{Q}$. Moreover in the worst case when $\mathbb{G}_p$ is strongly rooted at just one vertex and all vertices are neighbors of at least one common vertex, $||\lfloor F_p \rfloor|| = 1 - \frac{1}{n}$. It follows that the worst case convergence rate is

$$\max_{p \in \mathcal{Q}} ||\lfloor F_p \rfloor|| = 1 - \frac{1}{n} \tag{6.25}$$

Rooted Graphs

The proof of Theorem 6.1 depends crucially on the fact that the graphs encountered along a trajectory of (6.3) are all strongly rooted. It is natural to ask if this requirement can be relaxed and still have all agents' headings converge to a common value. The aim of this section is to show that this can indeed be accomplished. To do this we need to have a meaningful way of "combining" sequences of graphs so that only the combined graph need be strongly rooted, but not necessarily the individual graphs making up the combination. One possible notion of combination of a sequence $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ would be that graph in $\mathcal{G}$ whose arc set is the union of the arc sets of the graphs in the sequence. It turns out that because we are interested in *sequences* of graphs rather than mere *sets* of graphs, a simple union is not quite the appropriate

notion for our purposes because a union does not take into account the order in which the graphs are encountered along a trajectory. What is appropriate is a slightly more general notion which we now define.

Composition of Graphs

Let us agree to say that the *composition* of a directed graph $\mathbb{G}_{p_1} \in \bar{\mathcal{G}}$ with a directed graph $\mathbb{G}_{p_2} \in \bar{\mathcal{G}}$, written $\mathbb{G}_{p_2} \circ \mathbb{G}_{p_1}$, is the directed graph with vertex set $\{1, \ldots, n\}$ and arc set defined in such a way so that (i, j) is an arc of the composition just in case there is a vertex q such that (i, q) is an arc of $\mathbb{G}_{p_1}$ and (q, j) is an arc of $\mathbb{G}_{p_2}$. Thus (i, j) is an arc of $\mathbb{G}_{p_2} \circ \mathbb{G}_{p_1}$ if and only if i has an observer in $\mathbb{G}_{p_1}$ which is also a neighbor of j in $\mathbb{G}_{p_2}$. Note that $\bar{\mathcal{G}}$ is closed under composition and that composition is an associative binary operation; because of this, the definition extend unambiguously to any finite sequence of directed graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$.

If we focus exclusively on graphs in $\mathcal{G}$, more can be said. In this case the definition of composition implies that the arcs of $\mathbb{G}_{p_1}$ and $\mathbb{G}_{p_2}$ are arcs of $\mathbb{G}_{p_2} \circ \mathbb{G}_{p_1}$. The definition also implies in this case that if $\mathbb{G}_{p_1}$ has a directed path from i to k and $\mathbb{G}_{p_2}$ has a directed path from k to j, then $\mathbb{G}_{p_2} \circ \mathbb{G}_{p_1}$ has a directed path from i to j. Both of these implications are consequences of the requirement that the vertices of the graphs in $\mathcal{G}$ all have self arcs. Note in addition that $\mathcal{G}$ is closed under composition. It is worth emphasizing that the union of the arc sets of a sequence of graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ in $\mathcal{G}$ must be contained in the arc set of their composition. However the converse is not true in general and it is for this reason that composition rather than union proves to be the more useful concept for our purposes.

Suppose that $A_p = \big(a_{ij}(p)\big)$ and $A_q = \big(a_{ij}(q)\big)$ are the adjacency matrices of $\mathbb{G}_p \in \bar{\mathcal{G}}$ and $\mathbb{G}_q \in \bar{\mathcal{G}}$ respectively. Then the adjacency matrix of the composition $\mathbb{G}_q \circ \mathbb{G}_p$ must be the matrix obtained by replacing all non-zero elements in $A_p A_q$ with ones. This is because the ijth entry of $A_p A_q$, namely

$$\sum_{k=1}^{n} a_{ik}(p) a_{kj}(q),$$

will be non-zero just in case there is at least one value of k for which both $a_{ik}(p)$ and $a_{kj}(q)$ are non-zero. This of course is exactly the condition for the ijth element of the adjacency matrix of the composition $\mathbb{G}_q \circ \mathbb{G}_p$ to be non-zero. Note that if S_1 and S_2 are $n \times n$ stochastic matrices for which $\gamma(S_1) = \mathbb{G}_p$ and $\gamma(S_2) = \mathbb{G}_q$, then the matrix which results by replacing by ones, all non-zero entries in the stochastic matrix $S_2 S_1$, must be the transpose of the adjacency matrix of $\mathbb{G}_q \circ \mathbb{G}_p$. In view of the definition of $\gamma(\cdot)$, it therefore must be true that $\gamma(S_2 S_1) = \gamma(S_2) \circ \gamma(S_1)$. This obviously generalizes to finite products of stochastic matrices.

Lemma 6.3. *For any sequence of $n \times n$ stochastic matrices $S_1, S_2, \ldots, S_j$,*

$$\gamma(S_j \cdots S_1) = \gamma(S_j) \circ \cdots \circ \gamma(S_1),$$

Compositions of Rooted Graphs

We now give several different conditions under which the composition of a sequence of graphs is strongly rooted.

Proposition 6.3. *Suppose $n > 1$ and let $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}$ be a finite sequence of rooted graphs in $\mathcal{G}$.*

1. *If $m \geq n^2$, then $\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is strongly rooted.*
2. *If $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}$ are all rooted at v and $m \geq n-1$, then $\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is strongly rooted at v.*

The requirement that all the graphs in the sequence be rooted at a single vertex v is obviously more restrictive than the requirement that all the graphs be rooted, but not necessarily at the same vertex. The price for the less restrictive assumption, is that the bound on the number of graphs needed in the more general case is much higher than the bound given in the case which all the graphs are rooted at v. It is undoubtedly true that the bound n^2 for the more general case is too conservative. The more special case when all graphs share a common root is relevant to the leader follower version of the problem which will be discussed later in these notes. Proposition 6.3 will be proved in a moment.

Note that a strongly connected graph is the same as a graph which is rooted at every vertex and that a complete graph is the same as a graph which is strongly rooted at every vertex. In view of these observations and Proposition 6.3 we can state the following

Proposition 6.4. *Suppose $n > 1$ and let $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}$ be a finite sequence of strongly connected graphs in $\mathcal{G}$. If $m \geq n-1$, then $\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is complete.*

To prove Proposition 6.3 we will need some more ideas. We say that a vertex $v \in \mathcal{V}$ is a *observer* of a subset $\mathcal{S} \subset \mathcal{V}$ in a graph $\mathbb{G} \in \bar{\mathcal{G}}$, if v is an observer of at least one vertex in $\mathcal{S}$. By the *observer function* of a graph $\mathbb{G} \in \bar{\mathcal{G}}$, written $\alpha(\mathbb{G}, \cdot\,)$ we mean the function $\alpha(\mathbb{G}, \cdot\,) : 2^{\mathcal{V}} \to 2^{\mathcal{V}}$ which assigns to each subset $\mathcal{S} \subset \mathcal{V}$, the subset of vertices in $\mathcal{V}$ which are observers of $\mathcal{S}$ in $\mathbb{G}$. Thus $j \in \alpha(\mathbb{G}, i)$ just in case $(i, j) \in \mathcal{A}(\mathbb{G})$. Note that if $\mathbb{G}_p \in \bar{\mathcal{G}}$ and $\mathbb{G}_q$ in $\mathcal{G}$, then

$$\alpha(\mathbb{G}_p, \mathcal{S}) \subset \alpha(\mathbb{G}_q \circ \mathbb{G}_p, \mathcal{S}), \quad \mathcal{S} \in 2^{\mathcal{V}} \tag{6.26}$$

because $\mathbb{G}_q \in \mathcal{G}$ implies that the arcs in $\mathbb{G}_p$ are all arcs in $\mathbb{G}_q \circ \mathbb{G}_p$. Observer functions have the following important property.

Lemma 6.4. *For all $\mathbb{G}_p, \mathbb{G}_q \in \bar{\mathcal{G}}$ and any non-empty subset $\mathcal{S} \subset \mathcal{V}$,*

$$\alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, \mathcal{S})) = \alpha(\mathbb{G}_q \circ \mathbb{G}_p, \mathcal{S}) \tag{6.27}$$

Proof: Suppose first that $i \in \alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, \mathcal{S}))$. Then (j, i) is an arc in $\mathbb{G}_q$ for some $j \in \alpha(\mathbb{G}_p, \mathcal{S})$. Hence (k, j) is an arc in $\mathbb{G}_p$ for some $k \in \mathcal{S}$. In view of the definition of composition, (k, i) is an arc in $\mathbb{G}_q \circ \mathbb{G}_p$ so $i \in \alpha(\mathbb{G}_q \circ \mathbb{G}_p, \mathcal{S})$. Since this holds for all $i \in \mathcal{V}$, $\alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, \mathcal{S})) \subset \alpha(\mathbb{G}_q \circ \mathbb{G}_p, \mathcal{S})$.

For the reverse inclusion, fix $i \in \alpha(\mathbb{G}_q \circ \mathbb{G}_p, \mathcal{S})$ in which case (k, i) is an arc in $\mathbb{G}_q \circ \mathbb{G}_p$ for some $k \in \mathcal{S}$. By definition of composition, there exists an $j \in \mathcal{V}$ such that (k, j) is an arc in $\mathbb{G}_p$ and (j, i) is an arc in $\mathbb{G}_q$. Thus $j \in \alpha(\mathbb{G}_p, \mathcal{S})$. Therefore $i \in \alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, \mathcal{S}))$. Since this holds for all $i \in \mathcal{V}$, $\alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, \mathcal{S})) \supset \alpha(\mathbb{G}_q \circ \mathbb{G}_p, \mathcal{S})$. Therefore (6.27) is true. ■

To proceed, let us note that each subset $\mathcal{S} \subset \mathcal{V}$ induces a unique subgraph of $\mathbb{G}$ with vertex set $\mathcal{S}$ and arc set $\mathcal{A}$ consisting of those arcs (i, j) of $\mathbb{G}$ for which both i and j are vertices of $\mathcal{S}$. This together with the natural partial ordering of $\mathcal{V}$ by inclusion provides a corresponding partial ordering of $\bar{\mathcal{G}}$. Thus if $\mathcal{S}_1$ and $\mathcal{S}_2$ are subsets of $\mathcal{V}$ and $\mathcal{S}_1 \subset \mathcal{S}_2$, then $\mathbb{G}_1 \subset \mathbb{G}_2$ where for $i \in \{1, 2\}$, $\mathbb{G}_i$ is the subgraph of $\mathbb{G}$ induced by $\mathcal{S}_i$. For any $v \in \mathcal{V}$, there is a unique largest subgraph rooted at v, namely the graph induced by the vertex set $\mathcal{V}(v) = \{v\} \cup \alpha(\mathbb{G}, v) \cup \cdots \cup \alpha^{n-1}(\mathbb{G}, v)$ where $\alpha^i(\mathbb{G}, \cdot)$ denotes the composition of $\alpha(\mathbb{G}, \cdot)$ with itself i times. We call this graph, the *rooted graph generated by v*. It is clear that $\mathcal{V}(v)$ is the smallest $\alpha(\mathbb{G}, \cdot)$ - invariant subset of $\mathcal{V}$ which contains v.

The proof of Propositions 6.3 depends on the following lemma.

Lemma 6.5. *Let $\mathbb{G}_p$ and $\mathbb{G}_q$ be graphs in $\mathcal{G}$. If $\mathbb{G}_q$ is rooted at v and $\alpha(\mathbb{G}_p, v)$ is a strictly proper subset of $\mathcal{V}$, then $\alpha(\mathbb{G}_p, v)$ is also a strictly proper subset of $\alpha(\mathbb{G}_q \circ \mathbb{G}_p, v)$*

Proof of Lemma 6.5: In general $\alpha(\mathbb{G}_p, v) \subset \alpha(\mathbb{G}_q \circ \mathbb{G}_p, v)$ because of (6.26). Thus if $\alpha(\mathbb{G}_p, v)$ is not a strictly proper subset of $\alpha(\mathbb{G}_q \circ \mathbb{G}_p, v)$, then $\alpha(\mathbb{G}_p, v) = \alpha(\mathbb{G}_q \circ \mathbb{G}_p, v)$ so $\alpha(\mathbb{G}_q \circ \mathbb{G}_p, v) \subset \alpha(\mathbb{G}_p, v)$. In view of (6.27), $\alpha(\mathbb{G}_q \circ \mathbb{G}_p, v) = \alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, v))$. Therefore $\alpha(\mathbb{G}_q, \alpha(\mathbb{G}_p, v)) \subset \alpha(\mathbb{G}_p, v)$. Moreover, $v \in \alpha(\mathbb{G}_p, v)$ because v has a self-arc in $\mathbb{G}_p$. Thus $\alpha(\mathbb{G}_p, v)$ is a strictly proper subset of $\mathcal{V}$ which contains v and is $\alpha(\mathbb{G}_q, \cdot)$ - invariant. But this is impossible because $\mathbb{G}_q$ is rooted at v. ■

Proof of Proposition 6.3: Suppose $m \geq n^2$. In view of (6.26), $\mathcal{A}(\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}) \subset \mathcal{A}(\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1})$ for any positive integer $k \leq m$. Thus $\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ will be strongly rooted if there exists an integer $k \leq n^2$ such that $\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is strongly rooted. It will now be shown that such an integer exists.

If $\mathbb{G}_{p_1}$ is strongly rooted, set $k = 1$. If $\mathbb{G}_{p_1}$ is not strongly rooted, then let $i > 1$ be the least positive integer not exceeding n^2 for which $\mathbb{G}_{p_{i-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is not strongly rooted. If $i < n^2$, set $k = i$ in which case $\mathbb{G}_{p_k} \circ \cdots \circ \mathbb{G}_{p_1}$ is strongly rooted. Therefore suppose $i = n^2$ in which case $\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is not strongly

rooted for $j \in \{2,3,\dots,n^2\}$. Fix $j \in \{2,3,\dots,n^2\}$ and let v_j be any root of $\mathbb{G}_{p_j}$. Since $\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is not strongly rooted, $\alpha(\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v_j)$ is a strictly proper subset of $\mathcal{V}$. Hence by Lemma 6.5, $\alpha(\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v_j)$ is also a strictly proper subset of $\alpha(\mathbb{G}_{p_j} \circ \cdots \circ \mathbb{G}_{p_1}, v_j)$. Thus $\mathcal{A}(\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1})$ is a strictly proper subset of $\mathcal{A}(\mathbb{G}_{p_j} \circ \cdots \circ \mathbb{G}_{p_1})$. Since this holds for all $j \in \{2,3,\dots,n^2\}$ each containment in the ascending chain

$$\mathcal{A}(\mathbb{G}_{p_1}) \subset \mathcal{A}(\mathbb{G}_{p_2} \circ \mathbb{G}_{p_1}) \subset \cdots \subset \mathcal{A}(\mathbb{G}_{p_{n^2}} \circ \cdots \circ \mathbb{G}_{p_1})$$

is strict. Since $\mathcal{A}(\mathbb{G}_{p_1})$ must contain at least one arc, and there are at most n^2 arcs in any graph in $\mathcal{G}$, $\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ must be strongly rooted if $k = n^2$.

Now suppose that $m \geq n-1$ and $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \dots, \mathbb{G}_{p_m}$ are all rooted at v. In view of (6.26), $\mathcal{A}(\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}) \subset \mathcal{A}(\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1})$ for any positive integer $k \leq m$. Thus $\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ will be strongly rooted at v if there exists an integer $k \leq n-1$ such that

$$\alpha(\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v) = \mathcal{V} \tag{6.28}$$

It will now be shown that such an integer exists.

If $\alpha(\mathbb{G}_{p_1}, v) = \mathcal{V}$, set $k = 1$ in which case (6.28) clearly holds. If $\alpha(\mathbb{G}_{p_1}, v) \neq \mathcal{V}$, then let $i > 1$ be the least positive integer not exceeding $n-1$ for which $\alpha(\mathbb{G}_{p_{i-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v)$ is a strictly proper subset of $\mathcal{V}$. If $i < n-1$, set $k = i$ in which case (6.28) is clearly true. Therefore suppose $i = n-1$ in which case $\alpha(\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v)$ is a strictly proper subset of $\mathcal{V}$ for $j \in \{2,3,\dots,n-1\}$. Hence by Lemma 6.5, $\alpha(\mathbb{G}_{p_{j-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v)$ is also a strictly proper subset of $\alpha(\mathbb{G}_{p_j} \circ \cdots \circ \mathbb{G}_{p_1}, v)$ for $j \in \{2,3,\dots,n-1\}$. In view of this and (6.26), each containment in the ascending chain

$$\alpha(\mathbb{G}_{p_1}, v) \subset \alpha(\mathbb{G}_{p_2} \circ \mathbb{G}_{p_1}, v) \subset \cdots \subset \alpha(\mathbb{G}_{p_{n-1}} \circ \cdots \circ \mathbb{G}_{p_1}, v)$$

is strict. Since $\alpha(\mathbb{G}_{p_1}, v)$ has at least two vertices in it, and there are n vertices in $\mathcal{V}$, (6.28) must hold with $k = n-1$. ■

Proposition 6.3 implies that every sufficiently long composition of graphs from a given subset $\widehat{\mathcal{G}} \subset \mathcal{G}$ will be strongly rooted if each graph in $\widehat{\mathcal{G}}$ is rooted. The converse is also true. To understand why, suppose that it is not in which case there would have to be a graph $\mathbb{G} \in \widehat{\mathcal{G}}$, which is not rooted but for which $\mathbb{G}^m$ is strongly rooted for m sufficiently large where $\mathbb{G}^m$ is the m-fold composition of $\mathbb{G}$ with itself. Thus $\alpha(\mathbb{G}^m, v) = \mathcal{V}$ where v is a root of $\mathbb{G}^m$. But via repeated application of (6.27), $\alpha(\mathbb{G}^m, v) = \alpha^m(\mathbb{G}, v)$ where $\alpha^m(\mathbb{G}, \cdot)$ is the m-fold composition of $\alpha(\mathbb{G}, \cdot)$ with itself. Thus $\alpha^m(\mathbb{G}, v) = \mathcal{V}$. But this can only occur if $\mathbb{G}$ is rooted at v because $\alpha^m(\mathbb{G}, v)$ is the set of vertices reachable from v along paths of length m. Since this is a contradiction, $\mathbb{G}$ cannot be rooted. We summarize.

Proposition 6.5. *Every possible sufficiently long composition of graphs from a given subset $\widehat{\mathcal{G}} \subset \mathcal{G}$ is strongly rooted, if and only if every graph in $\widehat{\mathcal{G}}$ is rooted.*

Sarymsakov Graphs

In the sequel we say that a vertex $v \in \mathcal{V}$ is a *neighbor* of a subset $\mathcal{S} \subset \mathcal{V}$ in a graph $\mathbb{G} \in \bar{\mathcal{G}}$, if v is a neighbor of at least one vertex in $\mathcal{S}$. By a *Sarymsakov Graph* is meant a graph $\mathbb{G} \in \bar{\mathcal{G}}$ with the property that for each pair of non-empty subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ in $\mathcal{V}$ which have no neighbors in common, $\mathcal{S}_1 \cup \mathcal{S}_2$ contains a smaller number of vertices than does the set of neighbors of $\mathcal{S}_1 \cup \mathcal{S}_2$. Such seemingly obscure graphs are so named because they are the graphs of an important class of non-negative matrices studied by Sarymsakov in [29]. In the sequel we will prove that Sarymsakov graphs are in fact rooted graphs. We will also prove that the class of rooted graphs we are primarily interested in, namely those in $\mathcal{G}$, are Sarymsakov graphs.

It is possible to characterize Sarymsakov graph a little more concisely using the following concept. By the *neighbor function* of a graph $\mathbb{G} \in \bar{\mathcal{G}}$, written $\beta(\mathbb{G}, \cdot\,)$, we mean the function $\beta(\mathbb{G}, \cdot\,) : 2^{\mathcal{V}} \to 2^{\mathcal{V}}$ which assigns to each subset $\mathcal{S} \subset \mathcal{V}$, the subset of vertices in $\mathcal{V}$ which are neighbors of $\mathcal{S}$ in $\mathbb{G}$. Thus in terms of β, a Sarymsakov graph is a graph $\mathbb{G} \in \bar{\mathcal{G}}$ with the property that for each pair of non-empty subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ in $\mathcal{V}$ which have no neighbors in common, $\mathcal{S}_1 \cup \mathcal{S}_2$ contains less vertices than does the set $\beta(\mathbb{G}, \mathcal{S}_1 \cup \mathcal{S}_2)$. Note that if $\mathbb{G} \in \mathcal{G}$, the requirement that $\mathcal{S}_1 \cup \mathcal{S}_2$ contain less vertices than $\beta(\mathbb{G}, \mathcal{S}_1 \cup \mathcal{S}_2)$ simplifies to the equivalent requirement that $\mathcal{S}_1 \cup \mathcal{S}_2$ be a strictly proper subset of $\beta(\mathbb{G}, \mathcal{S}_1 \cup \mathcal{S}_2)$. This is because every vertex in $\mathbb{G}$ is a neighbor of itself if $\mathbb{G} \in \mathcal{G}$.

Proposition 6.6.

1. *Each Sarymsakov graph in $\bar{\mathcal{G}}$ is rooted.*
2. *Each rooted graph in $\mathcal{G}$ is a Sarymsakov graph.*

It follows that if we restrict attention exclusively to graphs in $\mathcal{G}$, then rooted graphs and Sarymsakov graphs are one and the same.

In the sequel $\beta^m(\mathbb{G}, \cdot)$ denotes the m-fold composition of $\beta(\mathbb{G}, \cdot)$ with itself. The proof of Proposition 6.6 depends on the following ideas.

Lemma 6.6. *Let $\mathbb{G} \in \bar{\mathcal{G}}$ be a Sarymsakov graph. Let $\mathcal{S}$ be a non-empty subset of $\mathcal{V}$ such that $\beta(\mathbb{G}, \mathcal{S}) \subset \mathcal{S}$. Let v be any vertex in $\mathcal{V}$. Then there exists a non-negative integer $m \leq n$ such that $\beta^m(\mathbb{G}, v) \cap \mathcal{S}$ is non-empty.*

Proof: If $v \in \mathcal{S}$, set $m = 0$. Suppose next that $v \notin \mathcal{S}$. Set $\mathcal{T} = \{v\} \cup \beta(\mathbb{G}, v) \cup \cdots \cup \beta^{n-1}(\mathbb{G}, v)$ and note that $\beta^n(\mathbb{G}, v) \subset \mathcal{T}$. Since $\beta(\mathbb{G}, \mathcal{T}) = \beta(\mathbb{G}, v) \cup \beta^2(\mathbb{G}, v) \cup \cdots \cup \beta^n(\mathbb{G}, v)$, it must be true that $\beta(\mathbb{G}, \mathcal{T}) \subset \mathcal{T}$. Therefore

$$\beta(\mathbb{G}, \mathcal{T} \cup \mathcal{S}) \subset \mathcal{T} \cup \mathcal{S} \tag{6.29}$$

Suppose $\beta(\mathbb{G}, \mathcal{T}) \cap \beta(\mathbb{G}, \mathcal{S})$ is empty. Then because $\mathbb{G}$ is a Sarymsakov graph, $\mathcal{T} \cup \mathcal{S}$ contains fewer vertices than $\beta(\mathbb{G}, \mathcal{T} \cup \mathcal{S})$. This contradicts (6.29) so $\beta(\mathbb{G}, \mathcal{T}) \cap \beta(\mathbb{G}, \mathcal{S})$ is not empty. In view of the fact that $\beta(\mathbb{G}, \mathcal{T}) = \beta(\mathbb{G}, v) \cup \beta^2(\mathbb{G}, v) \cup \cdots \cup \beta^n(\mathbb{G}, v)$ it must therefore be true for some positive integer $m \leq n$, that $\beta^m(\mathbb{G}, v) \cap \mathcal{S}$ is non-empty. ■

Lemma 6.7. *Let $\mathbb{G} \in \bar{\mathcal{G}}$ be rooted at r. Each non-empty subset $\mathcal{S} \subset \mathcal{V}$ not containing r is a strictly proper subset of $\mathcal{S} \cup \beta(\mathbb{G}, \mathcal{S})$.*

Proof of Lemma 6.7: Let $\mathcal{S} \subset \mathcal{V}$ be non-empty and not containing r. Pick $v \in \mathcal{S}$. Since $\mathbb{G}$ is rooted at r, there must be a path in $\mathbb{G}$ from r to v. Since $r \notin \mathcal{S}$ there must be a vertex $x \in \mathcal{S}$ which has a neighbor which is not in $\mathcal{S}$. Thus there is a vertex $y \in \beta(\mathbb{G}, \mathcal{S})$ which is not in $\mathcal{S}$. This implies that $\mathcal{S}$ is a strictly proper subset of $\mathcal{S} \cup \beta(\mathbb{G}, \mathcal{S})$. ■

By a *maximal rooted subgraph of* $\mathbb{G}$ we mean a subgraph $\mathbb{G}^*$ of $\mathbb{G}$ which is rooted and which is not contained in any rooted subgraph of $\mathbb{G}$ other than itself. Graphs in $\bar{\mathcal{G}}$ may have one or more maximal rooted subgraphs. Clearly $\mathbb{G}^* = \mathbb{G}$ just in case $\mathbb{G}$ is rooted. Note that if $\widehat{\mathcal{R}}$ is the set of all roots of a maximal rooted subgraph $\widehat{\mathbb{G}}$, then $\beta(\mathbb{G}, \widehat{\mathcal{R}}) \subset \widehat{\mathcal{R}}$. For if this were not so, then it would be possible to find a vertex $x \in \beta(\mathbb{G}, \widehat{\mathcal{R}})$ which is not in $\widehat{\mathcal{R}}$. This would imply the existence of a path from x to some root $\widehat{v} \in \widehat{\mathcal{R}}$; consequently the graph induced by the set of vertices along this path together with $\widehat{\mathcal{R}}$ would be rooted at $x \notin \widehat{\mathcal{R}}$ and would contain $\widehat{\mathbb{G}}$ as a strictly proper subgraph. But this contradicts the hypothesis that $\widehat{\mathbb{G}}$ is maximal. Therefore $\beta(\mathbb{G}, \widehat{\mathcal{R}}) \subset \widehat{\mathcal{R}}$. Now suppose that $\widehat{\mathbb{G}}$ is any rooted subgraph in $\mathcal{G}$. Suppose that $\widehat{\mathbb{G}}$'s set of roots $\widehat{\mathcal{R}}$ satisfies $\beta(\mathbb{G}, \widehat{\mathcal{R}}) \subset \widehat{\mathcal{R}}$. We claim that $\widehat{\mathbb{G}}$ must then be maximal. For if this were not so, there would have to be a rooted graph $\mathbb{G}^*$ containing $\widehat{\mathbb{G}}$ as a strictly proper subset. This in turn would imply the existence of a path from a root x^* of $\mathbb{G}^*$ to a root v of $\widehat{\mathbb{G}}$; consequently $x^* \in \beta^i(\mathbb{G}, \widehat{\mathcal{R}})$ for some $i \geq 1$. But this is impossible because $\widehat{\mathcal{R}}$ is $\beta(\mathbb{G}, \cdot\,)$ invariant. Thus $\widehat{\mathbb{G}}$ is maximal. We summarize.

Lemma 6.8. *A rooted subgraph of a graph $\mathbb{G}$ generated by any vertex $v \in \mathcal{V}$ is maximal if and only if its set of roots is $\beta(\mathbb{G}, \cdot\,)$ invariant.*

Proof of Proposition 6.6: Write $\beta(\cdot)$ for $\beta(\mathbb{G}, \cdot)$. To prove the assertion 1, pick $\mathbb{G} \in \bar{\mathcal{G}}$. Let $\mathbb{G}^*$ be any maximal rooted subgraph of $\mathbb{G}$ and write $\mathcal{R}$ for its root set; in view of Lemma 6.8, $\beta(\mathcal{R}) \subset \mathcal{R}$. Pick any $v \in \mathcal{V}$. Then by Lemma 6.6, for some positive integer $m \leq n$, $\beta^m(v) \cap \mathcal{R}$ is non-empty. Pick $z \in \beta^m(v) \cap \mathcal{R}$. Then there is a path from z to v and z is a root of $\mathbb{G}^*$. But $\mathbb{G}^*$ is maximal so v must be a vertex of $\mathbb{G}^*$. Therefore every vertex of $\mathbb{G}$ is a vertex of $\mathbb{G}^*$ which implies that $\mathbb{G}$ is rooted.

To prove assertion 2, let $\mathbb{G} \in \mathcal{G}$ be rooted at r. Pick any two non-empty subsets $\mathcal{S}_1, \mathcal{S}_2$ of $\mathcal{V}$ which have no neighbors in common. If $r \notin \mathcal{S}_1 \cup \mathcal{S}_2$, then $\mathcal{S}_1 \cup \mathcal{S}_2$ must be a strictly proper subset of $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \beta(\mathcal{S}_1 \cup \mathcal{S}_2)$ because of Lemma 6.7.

Suppose next that $r \in \mathcal{S}_1 \cup \mathcal{S}_2$. Since $\mathbb{G} \in \mathcal{G}$, $\mathcal{S}_i \subset \beta(\mathcal{S}_i)$, $i \in \{1, 2\}$. Thus $\mathcal{S}_1$ and $\mathcal{S}_2$ must be disjoint because $\beta(\mathcal{S}_1)$ and $\beta(\mathcal{S}_2)$ are. Therefore r must be in either $\mathcal{S}_1$ or $\mathcal{S}_2$ but not both. Suppose that $r \notin \mathcal{S}_1$. Then $\mathcal{S}_1$ must be a strictly proper subset of $\beta(\mathcal{S}_1)$ because of Lemma 6.7. Since $\beta(\mathcal{S}_1)$ and $\beta(\mathcal{S}_2)$ are disjoint, $\mathcal{S}_1 \cup \mathcal{S}_2$ must be a strictly proper subset of $\beta(\mathcal{S}_1 \cup \mathcal{S}_2)$. By the same

reasoning, $\mathcal{S}_1 \cup \mathcal{S}_2$ must be a strictly proper subset of $\beta(\mathcal{S}_1 \cup \mathcal{S}_2)$ if $r \notin \mathcal{S}_2$. Thus in conclusion $\mathcal{S}_1 \cup \mathcal{S}_2$ must be a strictly proper subset of $\beta(\mathcal{S}_1 \cup \mathcal{S}_2)$ whether r is in $\mathcal{S}_1 \cup \mathcal{S}_2$ or not. Since this conclusion holds for all such $\mathcal{S}_1$ and $\mathcal{S}_2$ and $\mathbb{G} \in \mathcal{G}$, $\mathbb{G}$ must be a Sarymsakov graph. ■

Neighbor-Shared Graphs

There is a different assumption which one can make about a sequence of graphs from $\bar{\mathcal{G}}$ which also insures that the sequence's composition is strongly rooted. For this we need the concept of a "neighbor-shared graph." Let us call $\mathbb{G} \in \bar{\mathcal{G}}$ *neighbor shared* if each set of 2 distinct vertices share a common neighbor. Suppose that $\mathbb{G}$ is neighbor shared. Then each pair of vertices is clearly reachable from a single vertex. Similarly each three vertices are reachable from paths starting at one of two vertices. Continuing this reasoning it is clear that each of the graph's n vertices is reachable from paths starting at vertices in some set $\mathcal{V}_{n-1}$ of $n-1$ vertices. By the same reasoning, each vertex in $\mathcal{V}_{n-1}$ is reachable from paths starting at vertices in some set $\mathcal{V}_{n-2}$ of $n-2$ vertices. Thus each of the graph's n vertices is reachable from paths starting at vertices in a set of $n-2$ vertices, namely the set $\mathcal{V}_{n-2}$. Continuing this argument we eventually arrive at the conclusion that each of the graph's n vertices is reachable from paths starting at a single vertex, namely the one vertex in the set $\mathcal{V}_1$. We have proved the following.

Lemma 6.9. *Each neighbor-shared graph in $\bar{\mathcal{G}}$ is rooted.*

It is worth noting that although shared neighbor graphs are rooted, the converse is not necessarily true. The reader may wish to construct a three vertex example which illustrates this. Although rooted graphs in $\mathcal{G}$ need not be neighbor shared, it turns out that the composition of any $n-1$ rooted graphs in $\mathcal{G}$ is.

Proposition 6.7. *The composition of any set of $m \geq n-1$ rooted graphs in $\mathcal{G}$ is neighbor shared.*

To prove Proposition 6.7 we need some more ideas. By the *reverse graph* of $\mathbb{G} \in \bar{\mathcal{G}}$, written $\mathbb{G}'$ is meant the graph in $\bar{\mathcal{G}}$ which results when the directions of all arcs in $\mathbb{G}$ are reversed. It is clear that $\mathcal{G}$ is closed under the reverse operation and that if A is the adjacency matrix of $\mathbb{G}$, then A' is the adjacency matrix of $\mathbb{G}'$. It is also clear that $(\mathbb{G}_p \circ \mathbb{G}_q)' = \mathbb{G}'_q \circ \mathbb{G}'_p, \quad p, q \in \bar{\mathcal{P}}$, and that

$$\alpha(\mathbb{G}', \mathcal{S}) = \beta(\mathbb{G}, \mathcal{S}), \quad \mathcal{S} \in 2^{\mathcal{V}} \tag{6.30}$$

Lemma 6.10. *For all $\mathbb{G}_p, \mathbb{G}_q \in \bar{\mathcal{G}}$ and any non-empty subset $\mathcal{S} \subset \mathcal{V}$,*

$$\beta(\mathbb{G}_q, \beta(\mathbb{G}_p, \mathcal{S})) = \beta(\mathbb{G}_p \circ \mathbb{G}_q, \mathcal{S}). \tag{6.31}$$

Proof of Lemma 6.10: In view of (6.27), $\alpha(\mathbb{G}'_p, \alpha(\mathbb{G}'_q, \mathcal{S})) = \alpha(\mathbb{G}'_p \circ \mathbb{G}'_q, \mathcal{S})$. But $\mathbb{G}'_p \circ \mathbb{G}'_q = (\mathbb{G}_q \circ \mathbb{G}_p)'$ so $\alpha(\mathbb{G}'_p, \alpha(\mathbb{G}'_q, \mathcal{S})) = \alpha((\mathbb{G}_q \circ \mathbb{G}_p)', \mathcal{S})$. Therefore $\beta(\mathbb{G}_p, \beta(\mathbb{G}_q, \mathcal{S})) = \beta(\mathbb{G}_q \circ \mathbb{G}_p), \mathcal{S})$ because of (6.30). ■

Lemma 6.11. *Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be rooted graphs in $\mathcal{G}$. If u and v are distinct vertices in $\mathcal{V}$ for which*

$$\beta(\mathbb{G}_2, \{u, v\}) = \beta(\mathbb{G}_2 \circ \mathbb{G}_1, \{u, v\}) \tag{6.32}$$

then u and v have a common neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$

Proof: $\beta(\mathbb{G}_2, u)$ and $\beta(\mathbb{G}_2, v)$ are non-empty because u and v are neighbors of themselves. Suppose u and v do not have a common neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$. Then $\beta(\mathbb{G}_2 \circ \mathbb{G}_1, u)$ and $\beta(\mathbb{G}_2 \circ \mathbb{G}_1, v)$ are disjoint. But $\beta(\mathbb{G}_2 \circ \mathbb{G}_1, u) = \beta(\mathbb{G}_1, \beta(\mathbb{G}_2, u))$ and $\beta(\mathbb{G}_2 \circ \mathbb{G}_1, v) = \beta(\mathbb{G}_1, \beta(\mathbb{G}_2, v))$ because of (6.31). Therefore $\beta(\mathbb{G}_1, \beta(\mathbb{G}_2, u))$ and $\beta(\mathbb{G}_1, \beta(\mathbb{G}_2, v))$ are disjoint. But $\mathbb{G}_1$ is rooted and thus a Sarymsakov graph because of Proposition 6.6. Thus $\beta(\mathbb{G}_2, \{u, v\})$ is a strictly proper subset of $\beta(\mathbb{G}_2, \{u, v\}) \cup \beta(\mathbb{G}_1, \beta(\mathbb{G}_2, \{u, v\})$. But $\beta(\mathbb{G}_2, \{u, v\}) \subset \beta(\mathbb{G}_1, \beta(\mathbb{G}_2, \{u, v\})$ because all vertices in $\mathbb{G}_2$ are neighbors of themselves and $\beta(\mathbb{G}_1, \beta(\mathbb{G}_2, \{u, v\}) = \beta(\mathbb{G}_2 \circ \mathbb{G}_1, \{u, v\})$ because of (6.31). Therefore $\beta(\mathbb{G}_2, \{u, v\})$ is a strictly proper subset of $\beta(\mathbb{G}_2 \circ \mathbb{G}_1, \{u, v\})$. This contradicts (6.32) so u and v have a common neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$. ■

Proof of Proposition 6.7: Let u and v be distinct vertices in $\mathcal{V}$. Let $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_{n-1}$ be a sequence of rooted graphs in $\mathcal{G}$. Since $\mathcal{A}(\mathbb{G}_{n-1} \circ \cdots \mathbb{G}_{n-i}) \subset \mathcal{A}(\mathbb{G}_{n-1} \circ \cdots \mathbb{G}_{n-(i+1)})$ for $i \in \{1, 2, \ldots, n-2\}$, it must be true that the $\mathbb{G}_i$ yield the ascending chain

$$\beta(\mathbb{G}_{n-1}, \{u, v\}) \subset \beta(\mathbb{G}_{n-1} \circ \mathbb{G}_{n-2}, \{u, v\}) \subset \cdots \subset \beta(\mathbb{G}_{n-1} \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{u, v\})$$

Because there are n vertices in $\mathcal{V}$, this chain must converge for some $i < n-1$ which means that

$$\beta(\mathbb{G}_{n-1} \circ \cdots \circ \mathbb{G}_{n-i}, \{u, v\}) = \beta(\mathbb{G}_{n-1} \circ \cdots \circ \mathbb{G}_{n-i} \circ \mathbb{G}_{n-(i+1)}, \{u, v\})$$

This and Lemma 6.11 imply that u and v have a common neighbor in $\mathbb{G}_{n-1} \circ \cdots \circ \mathbb{G}_{n-i}$ and thus in $\mathbb{G}_{n-1} \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1$. Since this is true for all distinct u and v, $\mathbb{G}_{n-1} \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1$ is a neighbor shared graph. ■

If we restrict attention to those rooted graphs in $\mathcal{G}$ which are strongly connected, we can obtain a neighbor-shared graph by composing a smaller number of rooted graphs that claimed in Proposition 6.7.

Proposition 6.8. *Let k be the integer quotient of n divided by 2. The composition of any set of $m \geq k$ strongly connected graphs in $\mathcal{G}$ is neighbor shared.*

Proof of Proposition 6.8: Let $k < n$ be a positive integer and let v be any vertex in $\mathcal{V}$. Let $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_k$ be a sequence of strongly connected

graphs in $\mathcal{G}$. Since $k < n$ and $\mathcal{A}(\mathbb{G}_k \circ \cdots \mathbb{G}_{k-i}) \subset \mathcal{A}(\mathbb{G}_{k-1} \circ \cdots \mathbb{G}_{k-(i+1)})$ for $i \in \{1, 2, \ldots, k-1\}$, it must be true that the $\mathbb{G}_i$ yield the ascending chain

$$\{v\} \subset \beta(\mathbb{G}_k, \{v\}) \subset \beta(\mathbb{G}_k \circ \mathbb{G}_{k-1}, \{v\}) \subset \cdots \subset \beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{v\})$$

Moreover, since any strongly connected graph is one in which every vertex is a root, it must also be true that the subsequence

$$\{v\} \subset \beta(\mathbb{G}_k, \{v\}) \subset \beta(\mathbb{G}_k \circ \mathbb{G}_{k-1}, \{v\}) \subset \cdots \subset \beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_{(i+1)} \circ \mathbb{G}_i, \{v\})$$

is strictly increasing where either $i = 1$ or $i > 1$ and $\beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_{(i+1)} \circ \mathbb{G}_i, \{v\}) = \mathcal{V}$. In either case this implies that $\beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{v\})$ contains at least $k+1$ vertices. Fix k as the integer quotient of $n+1$ divided by 2 in which case $2k \geq n-1$. Let v_1 and v_2 be any pair of distinct vertices in $\mathcal{V}$. Then there must be at least $k+1$ vertices in $\beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{v_1\})$ and $k+1$ vertices in $\beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{v_2\})$. But $2(k+1) > n$ so $\beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{v_1\})$ and $\beta(\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1, \{v_2\})$ must have at least one vertex in common. Since this is true for each pair of distinct vertices $v_1, v_2 \in \mathcal{V}$, $\mathbb{G}_k \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1$ must be neighbor-shared. ■

Lemma 6.9 and Proposition 6.3 imply that any composition of n^2 neighbor-shared graphs in $\mathcal{G}$ is strongly rooted. The following proposition asserts that the composition need only consist of $n-1$ neighbor-shared graphs and moreover that the graphs need only be in $\bar{\mathcal{G}}$ and not necessarily in $\mathcal{G}$.

Proposition 6.9. *The composition of any set of $m \geq n-1$ neighbor-shared graphs in $\bar{\mathcal{G}}$ is strongly rooted.*

To prove this proposition we need a few more ideas. For any integer $1 < k \leq n$, we say that a graph $\mathbb{G} \in \bar{\mathcal{G}}$ is *k-neighbor shared* if each set of k distinct vertices share a common neighbor. Thus a neighbor-shared graph and a 2 neighbor shared graph are one and the same. Clearly a n neighbor shared graph is strongly rooted at the common neighbor of all n vertices.

Lemma 6.12. *If $\mathbb{G}_p \in \bar{\mathcal{G}}$ is a neighbor-shared graph and $\mathbb{G}_q \in \bar{\mathcal{G}}$ is a k neighbor shared graph with $k < n$, then $\mathbb{G}_q \circ \mathbb{G}_p$ is a $(k+1)$ neighbor shared graph.*

Proof: Let $v_1, v_2, \ldots, v_{k+1}$ be any distinct vertices in $\mathcal{V}$. Since $\mathbb{G}_q$ is a k neighbor shared graph, the vertices $v_1, v_2, \ldots, v_k$ share a common neighbor u_1 in $\mathbb{G}_q$ and the vertices $v_2, v_2, \ldots, v_{k+1}$ share a common neighbor u_2 in $\mathbb{G}_q$ as well. Moreover, since $\mathbb{G}_p$ is a neighbor shared graph, u_1 and u_2 share a common neighbor w in $\mathbb{G}_p$. It follows from the definition of composition that $v_1, v_2, \ldots, v_k$ have w as a neighbor in $\mathbb{G}_q \circ \mathbb{G}_p$ as do $v_2, v_3, \ldots, v_{k+1}$. Therefore $v_1, v_2, \ldots, v_{k+1}$ have w as a neighbor in $\mathbb{G}_q \circ \mathbb{G}_p$. Since this must be true for any set of $k+1$ vertices in $\mathbb{G}_q \circ \mathbb{G}_p$, $\mathbb{G}_q \circ \mathbb{G}_p$ must be a $k+1$ neighbor shared graph as claimed. ■

Proof of Proposition 6.9: The preceding lemma implies that the composition of any two neighbor shared graphs is 3 neighbor shared. From this and

induction it follows that for $m < n$, the composition of m neighbor shared graphs is $m + 1$ neighbor shared. Thus the composition of $n - 1$ neighbor shared graphs is n neighbor shared and consequently strongly rooted. ■

In view of Proposition 6.9, we have the following slight improvement on Proposition 6.3.

Proposition 6.10. *The composition of any set of $m \geq (n-1)^2$ rooted graphs in $\mathcal{G}$ is strongly rooted.*

Convergence

We are now in a position to significantly relax the conditions under which the conclusion of Theorem 6.1 holds.

Theorem 6.2. *Let $\mathcal{Q}$ denote the subset of $\mathcal{P}$ consisting of those indices q for which $\mathbb{G}_q \in \mathcal{G}$ is rooted. Let $\theta(0)$ be fixed and let $\sigma : \{0, 1, 2, \ldots\} \to \mathcal{P}$ be a switching signal satisfying $\sigma(t) \in \mathcal{Q}$, $t \in \{0, 1, \ldots\}$. Then there is a constant steady state heading θ_{ss} depending only on $\theta(0)$ and σ for which*

$$\lim_{t\to\infty} \theta(t) = \theta_{ss}\mathbf{1} \tag{6.33}$$

where the limit is approached exponentially fast.

The theorem says that a unique heading is achieved asymptotically along any trajectory on which all neighbor graphs are rooted. The proof of Theorem 6.2 relies on the following generalization of Proposition 6.2.

Proposition 6.11. *Let $\mathcal{S}_r$ be any closed set of stochastic matrices in $\mathcal{S}$ whose graphs are all rooted. Then any product $S_j \ldots S_1$ of matrices from $\mathcal{S}_{sr}$ converges to $\mathbf{1}\lfloor \cdots S_j \cdots S_1 \rfloor$ exponentially fast as $j \to \infty$ at a rate no slower than λ, where λ is a non-negative constant depending on $\mathcal{S}_r$ and satisfying $\lambda < 1$.*

Proof of Proposition 6.11: Set $m = n^2$ and write $\mathcal{S}_r^m$ for the closed set of all products of stochastic matrices of the form $S_m S_{m-1} \cdots S_1$ where each $S_i \in \mathcal{S}_r$. By assumption, $\gamma(S)$ is rooted for $S \in \mathcal{S}$. In view of Proposition 6.3, $\gamma(S_m) \circ \cdots \gamma(S_1)$ is strongly rooted for every list of m matrices $\{S_1, S_2, \ldots, S_m\}$ from $\mathcal{S}_r$. But $\gamma(S_m) \circ \cdots \circ \gamma(S_1) = \gamma(S_m \cdots S_1)$ because of Lemma 6.3. Therefore $\gamma(S_m \cdots S_1)$ is strongly rooted for all products $S_m \cdots S_1 \in \mathcal{S}_r^m$.

Now any product $S_j \cdots S_1$ of matrices in $\mathcal{S}_r$ can be written as

$$S_j \cdots S_1 = \bar{S}(j)\bar{S}_k \cdots \bar{S}_1$$

where

$$\bar{S}_i = S_{im} \cdots S_{(i-1)m+1}, \quad 1 \leq i \leq k$$

is a product in $\mathcal{S}_r^m$,

$$\bar{S}(j) = S_j \cdots S_{(km+1)},$$

and k is the integer quotient of j divided by m. In view of Proposition 6.2, $\bar{S}_k \cdots \bar{S}_1$ must converge to $\mathbf{1}\lfloor \cdots \bar{S}_k \cdots \bar{S}_1 \rfloor$ exponentially fast as $k \to \infty$ at a rate no slower than $\bar{\lambda}$, where

$$\bar{\lambda} = \max_{\bar{S} \in \mathcal{S}_r^m} ||\lfloor \bar{S} \rceil||$$

But $\bar{S}(j)$ is a product of at most m stochastic matrices, so it is a bounded function of j. It follows that the product $S_j S_{j-1} \cdots S_1$ must converge to $\mathbf{1}\lfloor \cdots S_j \cdot S_1 \rfloor$ exponentially fast at a rate no slower than $\lambda = \bar{\lambda}^{\frac{1}{m}}$. ■

The proof of Proposition 6.11 can be applied to any closed subset $\mathcal{S}_{ns} \subset \mathcal{S}$ of stochastic matrices with neighbor shared graphs. In this case, one would define $m = n - 1$ because of Proposition 6.9.

Proof of Theorem 6.2: By definition, the graph $\mathbb{G}_p$ of each matrix F_p in the finite set $\{F_p : p \in \mathcal{Q}\}$ is rooted. By assumption, $F_{\sigma(t)} \in \{F_p : p \in \mathcal{Q}\}$, $t \geq 0$. In view of Proposition 6.11, the product $F_{\sigma(t)} \cdots F_{\sigma(0)}$ converges to $\mathbf{1}\lfloor \cdots F_{\sigma(t)} \cdots F_{\sigma(0)} \rfloor$ exponentially fast at a rate no slower than

$$\lambda = \{ \max_{F \in \mathcal{F}_r^m} ||\lfloor F \rceil||\}^{\frac{1}{m}}$$

where $m = n^2$ and $\mathcal{F}_r^m$ is the finite set of all m-term flocking matrix products of the form $F_{p_m} \cdots F_{p_1}$, $p_i \in \mathcal{Q}$. But it is clear from (6.3) that

$$\theta(t) = F_{\sigma(t-1)} \cdots F_{\sigma(1)} F_{\sigma(0)} \theta(0), \quad t \geq 1$$

Therefore (6.33) holds with $\theta_{ss} = \lfloor \cdots F_{\sigma(t)} \cdots F_{\sigma(0)} \rfloor \theta(0)$ and the convergence is exponential. ■

The proof of Theorem 6.2 also applies to the case when all of the $\mathbb{G}_{\sigma(t)}$, $t \geq 0$ are neighbor shared. In this case, one would define $m = n - 1$ because of Proposition 6.9.

Convergence Rates

It is possible to deduce an explicit convergence rate for the situation addressed in Theorem 6.2 [42]. To do this we need a few more ideas.

Scrambling Constants

Let S be an $n \times n$ stochastic matrix. Observe that for any non-negative n -vector x, the ith minus the jth entries of Sx can be written as

$$\sum_{k=1}^{n} (s_{ik} - s_{jk}) x_k = \sum_{k \in \mathcal{K}} (s_{ik} - s_{jk}) x_k + \sum_{k \in \bar{\mathcal{K}}} (s_{ik} - s_{jk}) x_k$$

where

$$\mathcal{K} = \{k : s_{ik} - s_{jk} \geq 0,\ k \in \{1, 2, \ldots, n\}\} \quad \text{and}$$
$$\bar{\mathcal{K}} = \{k : s_{ik} - s_{jk} < 0,\ k \in \{1, 2, \ldots, n\}\}$$

Therefore

$$\sum_{k=1}^{n}(s_{ik}-s_{jk})x_k \le \left(\sum_{k\in\mathcal{K}}(s_{ik}-s_{jk})\right)\lceil x\rceil + \left(\sum_{k\in\bar{\mathcal{K}}}(s_{ik}-s_{jk})\right)\lfloor x\rfloor$$

But

$$\sum_{k\in\mathcal{K}\cup\bar{\mathcal{K}}}(s_{ik}-s_{jk}) = 0$$

so

$$\sum_{k\in\bar{\mathcal{K}}}(s_{ik}-s_{jk}) = -\sum_{k\in\mathcal{K}}(s_{ik}-s_{jk})$$

Thus

$$\sum_{k=1}^{n}(s_{ik}-s_{jk})x_k \le \left(\sum_{k\in\mathcal{K}}(s_{ik}-s_{jk})\right)(\lceil x\rceil - \lfloor x\rfloor)$$

Now

$$\sum_{k\in\mathcal{K}}(s_{ik}-s_{jk}) = 1 - \sum_{k\in\bar{\mathcal{K}}}s_{ik} - \sum_{k\in\mathcal{K}}s_{jk}$$

because the row sums of S are all one. Moreover

$$\begin{aligned} s_{ik} &= \min\{s_{ik}, s_{jk}\}, \ \in \bar{\mathcal{K}} \\ s_{jk} &= \min\{s_{ik}, s_{jk}\}, \ \in \mathcal{K} \end{aligned}$$

so

$$\sum_{k\in\mathcal{K}}(s_{ik}-s_{jk}) = 1 - \sum_{k=1}^{n}\min\{s_{ik}, s_{jk}\}$$

It follows that

$$\sum_{k=1}^{n}(s_{ik}-s_{jk})x_k \le \left(1 - \sum_{k=1}^{n}\min\{s_{ik}, s_{jk}\}\right)(\lceil x\rceil - \lfloor x\rfloor)$$

Hence if we define

$$\mu(S) = \max_{i,j}\left(1 - \sum_{k=1}^{n}\min\{s_{ik}, s_{jk}\}\right) \tag{6.34}$$

then

$$\sum_{k=1}^{n}(s_{ik}-s_{jk})x_k \le \mu(S)(\lceil x\rceil - \lfloor x\rfloor)$$

Since this holds for all i, j, it must hold for that i and j for which

$$\sum_{k=1}^{n}s_{ik}x_k = \lceil Sx\rceil \qquad \text{and} \qquad \sum_{k=1}^{n}s_{jk}x_k = \lfloor Sx\rfloor$$

Therefore

$$\lceil Sx \rceil - \lfloor Sx \rfloor \leq \mu(S)(\lceil x \rceil - \lfloor x \rfloor) \tag{6.35}$$

Now let S_1 and S_2 be any two $n \times n$ stochastic matrices and let e_i be the ith unit n-vector. Then from (6.35),

$$\lceil S_2 S_1 e_i \rceil - \lfloor S_2 S_1 e_i \rfloor \leq \mu(S_2)(\lceil S_1 e_i \rceil - \lfloor S_1 e_1 \rfloor) \tag{6.36}$$

Meanwhile, from (6.11),

$$[\![S_2 S_1]\!] e_i = \mathbf{1}(\lceil S_2 S_1 \rceil - \lfloor S_2 S_1 \rfloor) e_i$$

and

$$[\![S_1]\!] e_i = \mathbf{1}(\lceil S_1 \rceil - \lfloor S_1 \rfloor) e_i$$

But for any non-negative matrix M, $\lceil M \rceil e_i = \lceil M e_i \rceil$ and $\lfloor M \rfloor e_i = \lfloor M e_i \rfloor$ so

$$[\![S_2 S_1]\!] e_i = \mathbf{1}(\lceil S_2 S_1 e_i \rceil - \lfloor S_2 S_1 e_i \rfloor)$$

and

$$[\![S_1]\!] e_i = \mathbf{1}(\lceil S_1 e_i \rceil - \lfloor S_1 e_i \rfloor)$$

¿From these expressions and (6.36) it follows that

$$[\![S_2 S_1]\!] e_i \leq \mu(s_2)[\![S_1]\!] e_i$$

Since this is true for all i, we arrive at the following fact.

Lemma 6.13. *For any two stochastic matrices in $\mathcal{S}$,*

$$[\![S_2 S_1]\!] \leq \mu(S_2)[\![S_1]\!] \tag{6.37}$$

where for any $n \times n$ stochastic matrix S,

$$\mu(S) = \max_{i,j} \left(1 - \sum_{k=1}^{n} \min\{s_{ik}, s_{jk}\} \right) \tag{6.38}$$

The quantity $\mu(S)$ has been widely studied before [29] and is know as the *scrambling constant* of the stochastic matrix S. Note that since the row sums of S all equal 1, $\mu(S)$ is non-negative. It is easy to see that $\mu(S) = 0$ just in case all the rows of S are equal. Let us note that for fixed i and j, the kth term in the sum appearing in (6.38) will be positive just in case both s_{ik} and s_{jk} are positive. It follows that the sum will be positive if and only if for at least one k, s_{ik} and s_{jk} are both positive. Thus $\mu(S) < 1$ if and only if for each distinct i and j, there is at least one k for which s_{ik} and s_{jk} are both positive. Matrices with this property have been widely studied and are called *scrambling matrices*. Thus a stochastic matrix S is a scrambling matrix if and only if $\mu(S) < 1$. It is easy to see that the definition of a scrambling matrix also implies that S is scrambling if and only if its graph $\gamma(S)$ is neighbor-shared.

The statement of Proposition 6.11 applies to the situation when instead of $\mathcal{S}_r$, one considers a closed subset $\mathcal{S}_{ns} \subset \mathcal{S}_r$ of stochastic matrices with neighbor shared graphs. In this case, the proof of Proposition 6.11 gives a worst case convergence rate bound of

$$\lambda = \left\{ \max_{\bar{S} \in \mathcal{S}_{ns}^m} ||\lfloor \bar{S} \rfloor|| \right\}^{\frac{1}{n-1}}$$

where $m = n-1$ and $\mathcal{S}_{ns}^m$ is the set of m term matrix products of the form $S_m \cdots S_1, \quad S_i \in \mathcal{S}_{ns}$. Armed with Lemma 6.13, one can do better.

Let $\mathcal{S}_{ns} \subset \mathcal{S}$ be a closed subset consisting of matrices whose graphs are all neighbor shared. Then the scrambling constant $\mu(S)$ defined in (6.38) satisfies $\mu(S) < 1,\ S \in \mathcal{S}_{ns}$ because each such S is a scrambling matrix. Let

$$\lambda = \max_{S \in \mathcal{S}_{ns}} \mu(S)$$

The $\lambda < 1$ because $\mathcal{S}_{ns}$ is closed and bounded and because $\mu(\cdot)$ is continuous. In view of Lemma 6.13,

$$||\lfloor S_2 S_1 \rfloor|| \leq \lambda ||\lfloor S_1 \rfloor||, \quad S_1, S_2 \in \mathcal{S}_{ns}$$

Hence by induction, for any sequence of matrices $S_1, S_2, \ldots$ in $\mathcal{S}_{ns}$

$$||\lfloor S_j \cdots S_1 \rfloor|| \leq \lambda^{j-1} ||\lfloor S_1 \rfloor||, \quad S_i \in \mathcal{S}_{ns}$$

But from (6.10), $\lfloor S \rfloor \leq \lfloor S \rfloor,\ \ S \in \mathcal{S}$, so $||\lfloor S \rfloor|| \leq ||\lfloor S \rfloor||,\ \ S \in \mathcal{S}$ Therefore for any sequence of stochastic matrices $S_1, S_2, \ldots$ with neighbor shared graphs

$$||\lfloor S_j \cdots S_1 \rfloor|| \leq \lambda^{j-1} ||\lfloor S_1 \rfloor|| \tag{6.39}$$

Therefore from Proposition 6.1, any such product $S_j \cdots S_1$ converges exponentially at a rate no slower than λ as $j \to \infty$.

Note that because of (6.22) and (6.10), the inequality in (6.39) applies to any sequence of stochastic matrices $S_1, S_2, \ldots$ for which $||\lfloor S_i \rfloor|| \leq \lambda$. Thus for example (6.39) applies to any sequence of stochastic matrices $S_1, S_2, \ldots$ whose graphs are strongly rooted provided the graphs in the sequence come from a compact set; in this case λ would be the maximum value of $||\lfloor S \rfloor||$ as S ranges over the set. Of course any such sequence is far more special then a sequence of stochastic matrices with neighbor-shared graphs since every strongly rooted graph is neighbor shared but the converse is generally not true.

Convergence Rates for Neighbor-Shared Graphs

Suppose that F_p is a flocking matrix for which $\mathbb{G}_p$ is neighbor shared. In view of the definition of a flocking matrix, any non-zero entry in F_p must be bounded below by $\frac{1}{n}$. Fix distinct i and j and suppose that k is a neighbor that i and j share. Then f_{ik} and f_{jk} are both non-zero so $\min\{f_{ik}, f_{jk}\} \geq \frac{1}{n}$. This implies that the sum in (6.38) must be bounded below by $\frac{1}{n}$ and consequently that $\mu(F_p) \leq 1 - \frac{1}{n}$.

Now let F_p be that flocking matrix whose graph $\mathbb{G}_p \in \mathcal{G}$ is such that vertex 1 has no neighbors other than itself, vertex 2 has every vertex as a neighbor, and vertices 3 through n have only themselves and agent 1 as neighbors. Since vertex 1 has no neighbors other than itself, $f_{i,k} = 0$ for all i and for $k > 1$. Thus for all i, j, it must be true that $\sum_{k=1}^{n} \min\{f_{ik}, f_{jk}\} = \min\{f_{i1}, f_{j1}\}$. Now vertex 2 has n neighbors, so $f_{2,1} = \frac{1}{n}$. Thus $\min\{f_{i1}, f_{j1}\}$ attains its lower bound of $\frac{1}{n}$ when either $i = 2$ or $j = 2$. It thus follows that with this F_p, $\mu(F_p)$ attains its upper bound of $1 - \frac{1}{n}$. We summarize.

Lemma 6.14. *Let $\mathcal{Q}$ be the set of indices in $\mathcal{P}$ for which $\mathbb{G}_p$ is neighbor shared. Then*

$$\max_{q \in \mathcal{Q}} \mu(F_q) = 1 - \frac{1}{n} \tag{6.40}$$

Lemma 6.14 can be used as follows. Let $\mathcal{Q}$ denote the set of $p \in \mathcal{P}$ for which $\mathbb{G}_p$ is neighbor shared. It is clear from the discussion at the end of the last section, that any product of flocking matrices $F_{p_j} \cdots F_{p_1},\ p_i \in \mathcal{Q}$ must converge at a rate no slower than

$$\lambda = \max_{q \in \mathcal{Q}} \mu(F_q)$$

Thus, in view of Lemma 6.14, $1 - \frac{1}{n}$ is a worst case bound on the rate of convergence of products of flocking matrices whose graphs are all neighbor shared. By way of comparison, $1 - \frac{1}{n}$ is a worst case bound if the flocking matrices in the product all have strongly rooted graphs {cf. (6.25)}. Of course the latter situation is far more special than the former, since strongly rooted graph are neighbor shared but not conversely.

Convergence Rates for Rooted Graphs

It is also possible to derive a worst case convergence rate for products of flocking matrices which have rooted rather than neighbor-shared graphs. As a first step towards this end we exploit the fact that for any $n \times n$ stochastic scrambling matrix S, the scrambling constant of $\mu(S)$ satisfies the inequality

$$\mu(S) \leq 1 - \phi(S) \tag{6.41}$$

where for any non-negative matrix M, $\phi(M)$ denote the smallest non-zero element of M. Assume that S is any scrambling matrix. Note that for any distinct i and j, there must be a k for which $\min\{s_{ik}, s_{jk}\}$ is non-zero and bounded above by $\phi(S)$. Thus

$$\sum_{k=1}^{n} \min\{s_{ik}, s_{jk}\} \geq \phi(S)$$

so

$$1 - \sum_{k=1}^{n} \min\{s_{ik}, s_{jk}\} \leq 1 - \phi(S)$$

But this holds for all distinct i and j. In view of the definition of $\mu(S)$ in (6.38), (6.41) must therefore be true.

We will also make use of the fact that for any two $n \times n$ stochastic matrices S_1 and S_2,

$$\phi(S_2 S_1) \geq \phi(S_2)\phi(S_1) \tag{6.42}$$

To prove that this is so note first that any stochastic matrix S can be written at $S = \phi(S)\bar{S}$ where $\bar{S}$ is a non-zero matrix whose non-zero entries are all bounded below by 1; moreover if $S = \widehat{S}\widehat{S}$ where $\widehat{\phi}(S)$ is a number and $\widehat{S}$ is also a non-zero matrix whose non-zero entries are all bounded below by 1, then $\phi(S) \geq \widehat{\phi}(S)$. Accordingly, write $S_i = \phi(S_i)\bar{S}_i,\ i \in \{1, 2\}$ where each $\bar{S}_i$ is a non-zero matrix whose non-zero entries are all bounded below by 1. Since $S_2 S_1 = \phi(S_2)\phi(S_1)\bar{S}_2\bar{S}_1$ and $S_2 S_1$ is non-zero, $\bar{S}_2\bar{S}_1$ must be non-zero as well. Moreover the nonzero entries of $\bar{S}_2\bar{S}_1$ must be bounded below by 1 because the product of any two $n \times n$ matrices with all non-zero entries bounded below by 1 must be a matrix with the same property. Therefore $\phi(S_2 S_1) \geq \phi(S_2)\phi(S_1)$ as claimed.

Suppose next that $\mathcal{S}_r$ is the set of all $n \times n$ stochastic matrices S which have rooted graphs $\gamma(S) \in \mathcal{G}$ and which satisfy $\phi(S) \geq b$ where b is a positive number smaller than 1. Thus for any set of $m = n - 1$ $S_i \in \mathcal{S}_r$,

$$\phi(S_m \cdots S_1) \geq b^m \tag{6.43}$$

because of (6.42). Now $\gamma(S_m \cdots S_1) = \gamma(S_m) \cdots \gamma(S_1)$. Moreover $\gamma(S_m \cdots S_1)$ will be neighbor shared if $m \geq n - 1$ because of Proposition 6.7. Therefore if $m \geq n - 1$, $S_m \cdots S_1$ is a scrambling matrix and

$$\mu(S_m \cdots S_1) \leq 1 - b^m \tag{6.44}$$

It turns out that this bound is actually attained if all the $S_i = S$, where S is a stochastic matrix of the form

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ b & 1-b & 0 & 0 & \cdots & 0 \\ 0 & b & 1-b & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1-b & 0 \\ 0 & 0 & 0 & 0 & b & 1-b \end{pmatrix} \tag{6.45}$$

with a graph which has no closed cycles other than the self-arcs at each of its vertices. To understand why this is so, note that the first row of S^m is $\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$ and the first element in the last row is b^m. This implies that $\min\{s_{11}, s_{n1}\} = b^m$, that

$$1 - \sum_{k=1}^{n} \min\{s_{1k}, s_{nk}\} \geq b^m,$$

and consequently that $\mu(S) \geq b^m$. We summarize.

Lemma 6.15. *Let b be a positive number less than 1 and let $m = n-1$. Let $\mathcal{S}_r^m$ denote the set of all m - term matrix products $S = S_m S_{m-1} \cdots S_1$ where each S_i is an $n \times n$ stochastic matrix with rooted graph $\gamma(S_i) \in \mathcal{G}$ and all-nonzero entries bounded below by b. Then*

$$\max_{S \in \mathcal{S}_r^m} \mu(S) = 1 - b^{(n-1)}$$

It is possible to apply at least part of the preceding to the case when the S_i are flocking matrices. Towards this end, let $m = n-1$ and let $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}$ be any sequence of m rooted graphs in $\mathcal{G}$ and let $F_{p_1}, \ldots, F_{p_m}$ be the sequence of flocking matrices associated with these graphs. Since each F_p is a flocking matrix, it must be true that $\phi(F_{p_i}) \geq \frac{1}{n}$, $i \in \{1, 2, \ldots, m\}$. Since the hypotheses leading up to (6.44) are satisfied,

$$\mu(F_{p_m} \cdots F_{p_1}) \leq 1 - \left(\frac{1}{n}\right)^{(n-1)} \tag{6.46}$$

Unfortunately, we cannot use the preceding reasoning to show that (6.46) holds with equality for some sequence of rooted graphs. This is because the matrix S in (6.45) is not a flocking matrix when $b = \frac{1}{n}$, except in the special case when $n = 2$. Nonetheless (6.46) can be used as follows to develop a convergence rate for products of flocking matrices whose graphs are all rooted. The development is very similar to that used in the proof of Proposition 6.11. Let $\mathcal{Q}$ denote the set of $p \in \mathcal{P}$ for which $\mathbb{G}_p$ is rooted and write $\mathcal{F}_r^m$ for the closed set of all products of flocking matrices of the form $F_{p_m} F_{p_{m-1}} \cdots F_{p_1}$ where each $p_i \in \mathcal{Q}$. In view of Proposition 6.7, $\mathbb{G}_{p_m} \circ \mathbb{G}_{p_{m-1}} \circ \cdots \mathbb{G}_{p_1}$ is neighbor shared for every list of m indices matrices $\{p_1, p_2, \ldots, p_m\}$ from $\mathcal{Q}$. Therefore (6.46) holds for every such list. Now for any sequence $p(1), p(2), \ldots, p(j)$ of indices in $\mathcal{Q}$, the corresponding product $F_{p(j)} \cdots F_{p(1)}$ of flocking matrices can be written as

$$F_{p(j)} \cdots F_{p(1)} = \bar{S}(j) \bar{S}_k \cdots \bar{S}_1$$

where

$$\bar{S}_i = F_{p(im)} \cdots F_{p((i-1)m+1)}, \quad 1 \leq i \leq k,$$
$$\bar{S}(j) = F_{p(j)} \cdots F_{p(km+1)},$$

and k is the integer quotient of j divided by m. In view of (6.46)

$$\mu(\bar{S}_i) \leq \bar{\lambda}, \quad i \in \{1, 2, \ldots, k\}$$

where

$$\bar{\lambda} = 1 - \left(\frac{1}{n}\right)^{(n-1)}$$

¿From this and the discussion at the end of the section on scrambling constants it is clear that $\bar{S}_k \cdots \bar{S}_1$ must converge to $\mathbf{1}\lfloor \cdots \bar{S}_k \cdots \bar{S}_1 \rfloor$ exponentially fast

as $k \to \infty$ at a rate no slower than $\bar{\lambda}$, But $\bar{S}(j)$ is a product of at most m stochastic matrices, so it is a bounded function of j. It follows that the product $F_{p(j)} \cdots F_{p(1)}$ must converge to $\mathbf{1}\lfloor F_{p(j)} \cdots F_{p(1)} \rfloor$ exponentially fast at a rate no slower than $\lambda = \bar{\lambda}^{\frac{1}{m}}$. We have proved the following corollary to Theorem 6.2.

Corollary 6.1. *Under the hypotheses of Theorem 6.2, convergence of $\theta(t)$ to $\theta_{ss}\mathbf{1}$ is exponential at a rate no slower than*

$$\lambda = \left\{1 - \left(\frac{1}{n}\right)^{(n-1)}\right\}^{\frac{1}{n-1}}$$

Convergence Rates for Strongly Connected Graphs

It is possible to develop results analogous to those in the last section for strongly connected graphs. Consider next the case when the $\mathbb{G}_p$ are all strongly connected. Suppose that $\mathcal{S}_{sc}$ is the set of all $n \times n$ stochastic matrices S which have strongly connected graphs $\gamma(S) \in \mathcal{G}$ and which satisfy $\phi(S) \geq b$ where b is a positive number smaller than 1. Let m denote the integer quotient of n divided by 2. Thus for any set of m stochastic matrices $S_i \in \mathcal{S}_{sc}$,

$$\phi(S_m \cdots S_1) \geq b^m \tag{6.47}$$

because of (6.42). Now $\gamma(S_m \cdots S_1) = \gamma(S_m) \cdots \gamma(S_1)$. Moreover $\gamma(S_m \cdots S_1)$ will be neighbor shared because of Proposition 6.8. Therefore $S_m \cdots S_1$ is a scrambling matrix and

$$\mu(S_m \cdots S_1) \leq 1 - b^m \tag{6.48}$$

Just as in the case of rooted graphs, it turns out that this bound is actually attained if all the $S_i = S$, where S is a stochastic matrix of the form

$$S = \begin{pmatrix} 1-b & 0 & 0 & 0 & \cdots & b \\ b & 1-b & 0 & 0 & \cdots & 0 \\ 0 & b & 1-b & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 1-b & 0 \\ 0 & 0 & 0 & 0 & b & 1-b \end{pmatrix} \tag{6.49}$$

with a graph consisting of one closed cycle containing all vertices plus self-arcs at each of its vertices. The reader may wish to verify that this is so by exploiting the structure of S^m.

Just as in the case of rooted graphs, it is possible to apply at least part of the preceding to the case when the S_i are flocking matrices. The development exactly parallels the rooted graph case and one obtains in the end the following corollary to Theorem 6.2.

Corollary 6.2. *Under the hypotheses of Theorem 6.2, and the additional assumption that σ takes values only in the subset of $\mathcal{Q}$ composed of those indices for which $\mathbb{G}_p$ is strongly connected, convergence of $\theta(t)$ to $\theta_{ss}\mathbf{1}$ is exponential at a rate no slower than*

$$\lambda = \left\{1 - \left(\frac{1}{n}\right)^m\right\}^{\frac{1}{m}}$$

where m is the integer quotient of n divided by 2.

Jointly Rooted Graphs

It is possible to relax still further the conditions under which the conclusion of Theorem 6.1 holds. Towards this end, let us agree to say that a finite sequence of directed graphs $\mathbb{G}_{p_1},\ \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ in $\mathcal{G}$ is *jointly rooted* if the composition $\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is rooted.

Note that since the arc set of any graph $\mathbb{G}_p, \mathbb{G}_q \in \mathcal{G}$ are contained in the arc set of any composed graph $\mathbb{G}_q \circ \mathbb{G} \circ \mathbb{G}_p$, $\mathbb{G} \in\in \mathcal{G}$, it must be true that if $\mathbb{G}_{p_1},\ \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ is a jointly rooted sequence, then so is $\mathbb{G}_q \circ \mathbb{G}_{p_1},\ \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}, \mathbb{G}_p$. In other words, a jointly rooted sequence of graphs in $\mathcal{G}$ remain jointly rooted if additional graphs from $\mathcal{G}$ are added to either end of the sequence.

There is an analogous concept for neighbor-shared graphs. We say that a finite sequence of directed graphs $\mathbb{G}_{p_1},\ \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ from $\mathcal{G}$ is *jointly neighbor-shared* if the composition $\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is a neighbor-shared graph. Jointly neighbor shared sequences of graphs remains jointly neighbor shared if additional graphs from $\mathcal{G}$ are added to either end of the sequence. The reason for this is the same as for the case of jointly rooted sequences. Although the discussion which follows is just for the case of jointly rooted graphs, the material covered extends in the obvious way to the case of jointly neighbor shared graphs.

In the sequel we will say that an infinite sequence of graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots,$ in $\mathcal{G}$ is *repeatedly jointly rooted* if there is a positive integer m for which each finite sequence $\mathbb{G}_{p_{m(k-1)+1}},\ \ldots, \mathbb{G}_{p_{mk}},\quad k \geq 1$ is jointly rooted. We are now in a position to generalize Proposition 6.11.

Proposition 6.12. *Let $\bar{\mathcal{S}}$ be any closed set of stochastic matrices in $\mathcal{S}$. Suppose that $S_1, S_2, \ldots$ is an infinite sequence of matrices from $\bar{\mathcal{S}}$ whose corresponding sequence of graphs $\gamma(S_1), \gamma(S_2), \ldots$ is repeatedly jointly rooted. Then the product $S_j \ldots S_1$ converges to $\mathbf{1}\lfloor \cdots S_j \cdots S_1 \rfloor$ exponentially fast as $j \to \infty$ at a rate no slower than λ, where $\lambda < 1$ is a non-negative constant depending on the sequence.*

Proof of Proposition 6.12: Since $\gamma(S_1), \gamma(S_2), \ldots$ is repeatedly jointly rooted, there is a finite integer m such that $\gamma(S_{m(k-1)+1}),\ \ldots, \gamma(S_{mk}),\quad k \geq 1$ is jointly rooted. For $k \geq 1$ define $\bar{S}_k = S_{mk} \cdots S_{m(k-1)+1}$. By Lemma 6.3,

$\gamma(S_{mk}\cdots S_{m(k-1)+1}) = \gamma(S_{mk})\circ\cdots\gamma(S_{m(k-1)+1}),\ k \geq 1$. Therefore $\gamma(\bar{S}_k)$ is rooted for $k \geq 1$. Note in addition that $\{\bar{S}_k, k \geq 1\}$ is a closed set because $\bar{S}$ is closed and m is finite. Thus by Proposition 6.11, the product $\bar{S}_k\cdots\bar{S}_1$ converges to $\mathbf{1}\lfloor\cdots\bar{S}_k\cdots\bar{S}_1\rfloor$ exponentially fast as $k\to\infty$ at a rate no slower than $\bar{\lambda}$, where $\bar{\lambda}$ is a non-negative constant depending on $\{\bar{S}_k, k \geq 1\}$ and satisfying $\bar{\lambda} < 1$.

Now the product $S_j\cdots S_1$ can be written as

$$S_j\cdots S_1 = \widehat{S}(j)\bar{S}_i\cdots\bar{S}_1$$

where

$$\widehat{S}(j) = S_j\cdots S_{(im+1)},$$

and i is the integer quotient of j divided by m. But $\widehat{S}(j)$ is a product of at most m stochastic matrices, so it is a bounded function of j. It follows that the product $S_jS_{j-1}\cdots S_1$ must converge to $\mathbf{1}\lfloor\cdots S_j\cdot S_1\rfloor$ exponentially fast at a rate no slower than $\lambda = \bar{\lambda}^{\frac{1}{m}}$. ■

We are now in a position to state our main result on leaderless coordination.

Theorem 6.3. *Let $\theta(0)$ be fixed and let $\sigma : [0,1,2,\ldots) \to \bar{\mathcal{P}}$ be a switching signal for which the infinite sequence of graphs $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ is repeatedly jointly rooted. Then there is a constant steady state heading θ_{ss}, depending only on $\theta(0)$ and σ, for which*

$$\lim_{t\to\infty}\theta(t) = \theta_{ss}\mathbf{1} \tag{6.50}$$

where the limit is approached exponentially fast.

Proof of Theorem 6.3: The set of flocking matrices $\mathcal{F} = \{F_p : p \in \mathcal{P}\}$ is finite and $\gamma(F_p) = \mathbb{G}_p,\ p \in \mathcal{P}$. Therefore the infinite sequence of matrices $F_{\sigma(0)}, F_{\sigma(1)}, \ldots$ come from a closed set and the infinite sequence of graphs $\gamma(F_{\sigma(0)}), \gamma(F_{\sigma(1)}), \ldots$ is repeatedly jointly rooted. It follows from Proposition 6.12 that the product $F_{\sigma(t)}\cdots F_{\sigma(1)}F_{\sigma(0)}$ converges to $\mathbf{1}\lfloor\cdots F_{\sigma(t)}\cdots F_{\sigma(1)}F_{\sigma(0)}\rfloor$ exponentially fast as $t\to\infty$ at a rate no slower than λ, where $\lambda < 1$ is a non-negative constant depending on the sequence. But it is clear from (6.3) that

$$\theta(t) = F_{\sigma(t-1)}\cdots F_{\sigma(1)}F_{\sigma(0)}\theta(0), \quad t \geq 1$$

Therefore (6.50) holds with $\theta_{ss} = \lfloor\cdots F_{\sigma(t)}\cdots F_{\sigma(0)}\rfloor\theta(0)$ and the convergence is exponential. ■

6.2 Symmetric Neighbor Relations

It is natural to call a graph in $\bar{\mathcal{G}}$ *symmetric* if for each pair of vertices i and j for which j is a neighbor of i, i is also a neighbor of j. Note that $\mathbb{G}$ is

symmetric if and only if its adjacency matrix is symmetric. It is worth noting that for symmetric graphs, the properties of rooted and rooted at v are both equivalent to the property that the graph is strongly connected. Within the class of symmetric graphs, neighbor-shared graphs and strongly rooted graphs are also strongly connected graphs but in neither case is the converse true. It is possible to represent a symmetric directed graph $\mathbb{G}$ with a simple undirected graph $\mathbb{G}^s$ in which each self-arc is replaced with an undirected edge and each pair of directed arcs (i, j) and (j, i) for distinct vertices is replaced with an undirected edge between i and j. Notions of strongly rooted and neighbor shared extend in the obvious way to unconnected graphs. An undirected graph is said to be *connected* if there is an undirected path between each pair of vertices. Thus a strongly connected, directed graph which is symmetric is in essence the same as a connected, undirected graph. Undirected graphs are applicable when the sensing radii r_i of all agents are the same. It was the symmetric version of the flocking problem which Vicsek addressed [25] and which was analyzed in [27] using undirected graphs.

Let $\bar{\mathcal{G}}_s$ and $\mathcal{G}_s$ denote the subsets of symmetric graphs in $\bar{\mathcal{G}}$ and $\mathcal{G}$ respectively. Simple examples show that neither $\bar{\mathcal{G}}_s$ nor $\mathcal{G}_s$ is closed under composition. In particular, composition of two symmetric directed graphs in $\bar{\mathcal{G}}$ is not typically symmetric. On the other hand the "union" is where by the *union* of $\mathbb{G}_p \in \bar{\mathcal{G}}$ and $\mathbb{G}_q \in \bar{\mathcal{G}}$ is meant that graph in $\bar{\mathcal{G}}$ whose arc set is the union of the arc sets of $\mathbb{G}_p$ and $\mathbb{G}_q$. It is clear that both $\bar{\mathcal{G}}^s$ and $\mathcal{G}^s$ are closed under the union operation. The union operation extends to undirected graphs in the obvious way. Specifically, the *union* of two undirected graphs with the same vertex set $\mathcal{V}$, is that graph whose vertex set is $\mathcal{V}$ and whose edge set is the union of the edge sets of the two graphs comprising the union. It is worth emphasizing that union and composition are really quite different operations. For example, as we've already seen with Proposition 6.4, the composition of any $n-1$ strongly connected graphs is complete, symmetric or not, is always complete. On the other hand, the union of $n-1$ $n-1$ strongly connected graphs is not necessarily complete. In terms of undirected graphs, it is simply not true that the union of $n-1$ undirected graphs with vertex set $\mathcal{V}$ is complete, even if each graph in the union has self-loops at each vertex. The root cause of the difference between union and composition stems from the fact that the union and composition of two graphs in $\bar{\mathcal{G}}$ have different arc sets – and in the case of graphs from $\mathcal{G}$, the arc set of the union is always contained in the arc set of the composition, but not conversely.

The development in [27] make use of the notion of a "jointly connected set of graphs." Specifically, a set of undirected graphs with vertex set $\mathcal{V}$ is *jointly connected* if the union of the graphs in the collection is a connected graph. The notion of jointly connected also applies to directed graphs in which case the collection is jointly connected if the union is strongly connected. In the sequel we will say that an infinite sequence of graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots$, in $\mathcal{G}$ is *repeatedly jointly connected* if there is a positive integer m for which each finite sequence

$\mathbb{G}_{p_{m(k-1)+1}}, \ldots, \mathbb{G}_{p_{mk}}, \quad k \geq 1$ is jointly connected. The main result of [27] is in essence as follows.

Theorem 6.4. *Let $\theta(0)$ be fixed and let $\sigma : [0, 1, 2, \ldots) \to \bar{\mathcal{P}}$ be a switching signal for which the infinite sequence of symmetric graphs $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ in $\mathcal{G}$ is repeatedly jointly connected. Then there is a constant steady state heading θ_{ss}, depending only on $\theta(0)$ and σ, for which*

$$\lim_{t \to \infty} \theta(t) = \theta_{ss}\mathbf{1} \tag{6.51}$$

where the limit is approached exponentially fast.

In view of Theorem 6.3, Theorem 6.4 also holds if the word "connected" is replaced with the word "rooted." The latter supposes that composition replaces union and that jointly rooted replaces jointly connected. Examples show that these modifications lead to a more general result because a jointly rooted sequence of graphs is always jointly connected but the converse is not necessarily true.

Generalization

It is possible to interpret the system we've been studying (6.3) as the closed-loop system which results when a suitably defined decentralized feedback law is applied to the n-agent heading model

$$\theta(t+1) = \theta(t) + u(t) \tag{6.52}$$

with open-loop control u. To end up with (6.3), u would have to be defined as

$$u(t) = -D^{-1}_{\sigma(t)} e(t) \tag{6.53}$$

where e is the *average heading error* vector

$$e(t) \stackrel{\Delta}{=} L_{\sigma(t)}\theta(t) \tag{6.54}$$

and, for each $p \in \mathcal{P}$, L_p is the matrix

$$L_p = D_p - A_p \tag{6.55}$$

known in graph theory as the *Laplacian* of $\mathbb{G}_p$ [43, 28]. It is easily verified that equations (6.52) to (6.55) do indeed define the system modelled by (6.3). We've elected to call e the average heading error because if $e(t) = 0$ at some time t, then the heading of each agent with neighbors at that time will equal the average of the headings of its neighbors.

In the present context, (6.53) can be viewed as a special case of a more general decentralized feedback control of the form

$$u(t) = -G^{-1}_{\sigma(t)} L_{\sigma(t)}\theta(t) \tag{6.56}$$

where for each $p \in \mathcal{P}$, G_p is a suitably defined, nonsingular diagonal matrix with ith diagonal element g_p^i. This, in turn, is an abbreviated description of a system of n individual agent control laws of the form

$$u_i(t) = -\frac{1}{g_i(t)} \left(\sum_{j \in \mathcal{N}_i(t)} \theta_j(t) \right), \quad i \in \{1, 2, \ldots, n\} \tag{6.57}$$

where for $i \in \{1, 2, \ldots, n\}$, $u_i(t)$ is the ith entry of $u(t)$ and $g_i(t) \stackrel{\Delta}{=} g^i_{\sigma(t)}$. Application of this control to (6.52) would result in the closed-loop system

$$\theta(t+1) = \theta(t) - G^{-1}_{\sigma(t)} L_{\sigma(t)} \theta(t) \tag{6.58}$$

Note that the form of (6.58) implies that if θ and σ were to converge to a constant values $\bar{\theta}$, and $\bar{\sigma}$ respectively, then $\bar{\theta}$ would automatically satisfy $L_{\bar{\sigma}} \bar{\theta} = 0$. This means that control (6.56) automatically forces each agent's heading to converge to the average of its neighbors, if agent headings were to converge at all. In other words, the choice of the G_p does not affect the requirement that each agent's heading equal the average of the headings of its neighbors, if there is convergence at all. In the sequel we will deal only with the case when the graphs $\mathbb{G}_p$ are all symmetric in which case L_p is symmetric as well.

The preceding suggests that there might be useful choices for the G_p alternative to those we've considered so far, which also lead to convergence. One such choice turns out to be

$$G_p = gI, \ p \in \mathcal{P} \tag{6.59}$$

where g is any number greater than n. Our aim is to show that with the G_p so defined, Theorem 6.2 continues to be valid. In sharp contrast with the proof technique used in the last section, convergence will be established here using a common quadratic Lyapunov function.

As before, we will use the model

$$\theta(t+1) = F_{\sigma(t)} \theta(t) \tag{6.60}$$

where, in view of the definition of the G_p in (6.59), the F_p are now symmetric matrices of the form

$$F_p = I - \frac{1}{g} L_p, \qquad p \in \mathcal{P} \tag{6.61}$$

To proceed we need to review a number of well known and easily verified properties of graph Laplacians relevant to the problem at hand. For this, let $\mathbb{G}$ be any given symmetric directed graph in $\mathcal{G}$. Let D be a diagonal matrix whose diagonal elements are the in-degrees of $\mathbb{G}$'s vertices and write A for $\mathbb{G}$'s adjacency matrix. Then, as noted before, the Laplacian of $\mathbb{G}$ is the symmetric matrix $L = D - A$. The definition of L clearly implies that $L\mathbf{1} = 0$. Thus

L must have an eigenvalue at zero and $\mathbf{1}$ must be an eigenvector for this eigenvalue. Surprisingly L is always a positive semidefinite matrix [28]. Thus L must have a real spectrum consisting of non-negative numbers and at least one of these numbers must be 0. It turns out that the number of connected components of $\mathbb{G}$ is exactly the same as the multiplicity of L's eigenvalue at 0 [28]. Thus $\mathbb{G}$ is a rooted or strongly connected graph just in case L has exactly one eigenvalue at 0. Note that the trace of L is the sum of the in-degrees of all vertices of $\mathbb{G}$. This number can never exceed $(n-1)n$ and can attain this high value only for a complete graph. In any event, this property implies that the maximum eigenvalue of L is never larger than $n(n-1)$. Actually the largest eigenvalue of L can never be larger than n [28]. This means that the eigenvalues of $\frac{1}{g}L$ must be smaller than 1 since $g > n$. From these properties it clearly follows that the eigenvalues of $(I - \frac{1}{g}L)$ must all be between 0 and 1, and that if $\mathbb{G}$ is strongly connected, then all will be strictly less than 1 except for one eigenvalue at 1 with eigenvector $\mathbf{1}$. Since each F_p is of the form $(I - \frac{1}{g}L)$, each F_p possesses all of these properties.

Let σ be a fixed switching signal with value $p_t \in \mathcal{Q}$ at time $t \geq 0$. What we'd like to do is to prove that as $i \to \infty$, the matrix product $F_{p_i}F_{p_{i-1}} \cdots F_{p_0}$ converges to $\mathbf{1}c$ for some row vector c. As noted near the beginning of section 6.1, this matrix product will so converge just in case

$$\lim_{i\to\infty} \tilde{F}_{p_i}\tilde{F}_{p_{i-1}} \cdots \tilde{F}_{p_0} = 0 \tag{6.62}$$

where as in section 6.1, $\tilde{F}_p$ is the unique solution to $PF_p = \tilde{F}_pP,\ p \in \mathcal{P}$ and P is any full rank $(n-1) \times n$ matrix satisfying $P\mathbf{1} = 0$. For simplicity and without loss of generality we shall henceforth assume that the rows of P form a basis for the orthogonal complement of the span of $\mathbf{e}$. This means that PP' equals the $(n-1)\times(n-1)$ identity $\tilde{I}$, that $\tilde{F}_p = PF_pP',\ \ p \in \mathcal{P}$, and thus that each $\tilde{F}_p$ is symmetric. Moreover, in view of (6.6) and the spectral properties of the $F_p,\ p \in \mathcal{Q}$, it is clear that each $\tilde{F}_p,\ p \in \mathcal{Q}$ must have a real spectrum lying strictly inside of the unit circle. This plus symmetry means that for each $p \in \mathcal{Q}$, $\tilde{F}_p - \tilde{I}$ is negative definite, that $\tilde{F}_p'\tilde{F}_p - \tilde{I}$ is negative definite and thus that $\tilde{I}$ is a common discrete-time Lyapunov matrix for all such $\tilde{F}_p$. Using this fact it is straight forward to prove that Theorem 6.2 holds for system (6.58) provided the G_p are defined as in (6.59) with $g > n$.

In general, each $\tilde{F}_p$ is a discrete-time stability matrix for which $\tilde{F}_p'\tilde{F}_p - \tilde{I}$ is negative definite only if $p \in \mathcal{Q}$. To craft a proof of Theorem 6.3 for the system described by (6.58) and (6.59), one needs to show that for each interval $[t_i, t_{i+1})$ on which $\{\mathbb{G}_{\sigma(t_{i+1}-1)}, \ldots \mathbb{G}_{\sigma(t_i+1)}, \mathbb{G}_{\sigma(t_i)}\}$ is a jointly rooted sequence of graphs, the product $\tilde{F}_{\sigma(t_{i+1}-1)} \cdots \tilde{F}_{\sigma(t_i+1)}\tilde{F}_{\sigma(t_i)}$ is a discrete-time stability matrix and

$$(\tilde{F}_{\sigma(t_{i+1}-1)} \cdots \tilde{F}_{\sigma(t_i+1)}\tilde{F}_{\sigma(t_i)})'(\tilde{F}_{\sigma(t_{i+1}-1)} \cdots \tilde{F}_{\sigma(t_i+1)}\tilde{F}_{\sigma(t_i)}) - \tilde{I}$$

is negative definite. This is a direct consequence of the following proposition.

Proposition 6.13. *If* $\{\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}\}$ *is a jointly rooted sequence of symmetric graphs, then*

$$(\tilde{F}_{p_1}\tilde{F}_{p_2}\cdots\tilde{F}_{p_m})'(\tilde{F}_{p_1}\tilde{F}_{p_2}\cdots\tilde{F}_{p_m}) - \tilde{I}$$

is a negative definite matrix.

In the light of Proposition 6.13, it is clear that the conclusion Theorem 6.3 is also valid for the system described by (6.58) and (6.59). A proof of this version of Theorem 6.3 will not be given.

To summarize, both the control defined by $u = -D^{-1}_{\sigma(t)}e(t)$ and the simplified control given by $u = -\frac{1}{g}e(t)$ achieve the same emergent behavior. While the latter is much easier to analyze than the former, it has the disadvantage of not being a true decentralized control because each agent must know an upper bound {i.e., g} on the total number of agents within the group. Whether or not this is really a disadvantage, of course depends on what the models are to be used for.

The proof of Proposition 6.13 depends on two lemmas. In the sequel, we state the lemmas, use them to prove Proposition 6.13, and then conclude this section with proofs of the lemmas themselves.

Lemma 6.16. *If* $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}$ *is a jointly rooted sequence of symmetric graphs in* $\mathcal{G}$ *with Laplacians* $L_{p_1}, L_{p_2}, \ldots, L_{p_m}$*, then*

$$\bigcap_{i=1}^{m} \text{kernel } L_{p_i} = \text{span } \{\mathbf{1}\} \tag{6.63}$$

Lemma 6.17. *Let* $M_1, M_2, \ldots, M_m$ *be a set of* $n \times n$ *real symmetric, matrices whose induced* 2*-norms are all less than or equal to* 1*. If*

$$\bigcap_{i=1}^{m} \text{kernel } (I - M_i) = 0 \tag{6.64}$$

then the induced 2*-norm of* $M_1M_2\cdots M_m$ *is less than* 1*.*

Proof of Proposition 6.13: The definition of the F_p in (6.61) implies that $I - F_p = \frac{1}{g}L_p$. Hence by Lemma 6.16 and the hypothesis that $\{\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}\}$ is a jointly rooted sequence,

$$\bigcap_{i=1}^{m} \text{kernel } (I - F_{p_i}) = \text{span } \{\mathbf{1}\} \tag{6.65}$$

We claim that

$$\bigcap_{i=1}^{m} \text{kernel } (\tilde{I} - \tilde{F}_{p_i}) = 0 \tag{6.66}$$

To establish this fact, let $\bar{x}$ be any vector such that $(\tilde{I} - \tilde{F}_{p_i})\bar{x} = 0,\ i \in \{1, 2, \ldots, m\}$. Since P has independent rows, there is a vector x such that $\bar{x} = Px$. But $P(I - F_{p_i}) = (\tilde{I} - \tilde{F}_{p_i})P$, so $P(I - F_{p_i})x = 0$. Hence $(I - F_{p_i})x = a_i\mathbf{1}$ for some number a_i. But $\mathbf{1}'(I - F_{p_i}) = \frac{1}{g}\mathbf{1}'L_{p_i} = 0$, so $a_i\mathbf{1}'\mathbf{1} = 0$. This implies that $a_i = 0$ and thus that $(I - F_{p_i})x = 0$. But this must be true for all $i \in \{1, 2, \ldots, m\}$. It follows from (6.65) that $x \in$ span $\{\mathbf{1}\}$ and, since $\bar{x} = Px$, that $\bar{x} = 0$. Therefore (6.66) is true.

As defined, the $\tilde{F}_p$ are all symmetric, positive semi-definite matrices with induced 2 - norms not exceeding 1. This and (6.66) imply that the family of matrices $\tilde{F}_{p_1}, \tilde{F}_{p_2}, \ldots, \tilde{F}_{p_m}$ satisfy the hypotheses of Lemma 6.17. It follows that Proposition 6.13 is true. ■

Proof of Lemma 6.16: In the sequel we write $L(\mathbb{G})$ for the Laplacian of a simple graph $\mathbb{G}$. By the *intersection* of a collection of graphs $\{\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}\}$ in $\mathcal{G}$, is meant that graph $\mathbb{G} \in \mathcal{G}$ with edge set equaling the intersection of the edge sets of all of the graphs in the collection. It follows at once from the definition of a Laplacian that

$$L(\mathbb{G}_p) + L(\mathbb{G}_q) = L(\mathbb{G}_p \cap \mathbb{G}_q) + L(\mathbb{G}_p \cup \mathbb{G}_q)$$

for all $p, q \in \mathcal{P}$. Repeated application of this identity to the set $\{\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}\}$ yields the relation

$$\sum_{i=1}^{m} L(\mathbb{G}_{p_i}) = L\left(\bigcup_{i=1}^{m} \mathbb{G}_{p_i}\right) + \sum_{i=1}^{m-1} L\left(\mathbb{G}_{p_{i+1}} \bigcap \left\{\bigcup_{j=1}^{i} \mathbb{G}_{p_j}\right\}\right) \tag{6.67}$$

which is valid for $m > 1$. Since all matrices in (6.67) are positive semi-definite, any vector x which makes the quadratic form $x'\{L(\mathbb{G}_{p_1}) + L(\mathbb{G}_{p_2}) + \cdots + L(\mathbb{G}_{p_m})\}x$ vanish, must also make the quadratic form $x'L(\mathbb{G}_{p_1} \cup \mathbb{G}_{p_2} \cup \cdots \cup \mathbb{G}_{p_m})x$ vanish. Since any vector in the kernel of each matrix $L(\mathbb{G}_{p_i})$ has this property, we can draw the following conclusion.

$$\bigcap_{i=1}^{m} \text{kernel } L(\mathbb{G}_{p_i}) \subset \text{kernel } L\left(\bigcup_{i=1}^{m} \mathbb{G}_{p_i}\right)$$

Suppose now that $\{\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_m}\}$ is a jointly rooted collection. Then the union $\mathbb{G}_{p_1} \cup \mathbb{G}_{p_2} \cup \cdots \cup \mathbb{G}_{p_m}$ is rooted so its Laplacian must have exactly span $\{\mathbf{1}\}$ for its kernel. Hence the intersection of the kernels of the $L(\mathbb{G}_{p_i})$ must be contained in span $\{\mathbf{1}\}$. But span $\{\mathbf{1}\}$ is contained in the kernel of each matrix $L(\mathbb{G}_{p_i})$ in the intersection and therefore in the intersection of the kernels of these matrices as well. It follows that (6.63) is true. ■

Proof of Lemma 6.17: In the sequel we write $|x|$ for the 2-norm of a real n-vector x and $|M|$ for the induced 2-norm of a real $n \times n$ matrix. Let $x \in \mathbb{R}^n$ be any real, non-zero n-vector. It is enough to show that

$$|M_1 M_2 \cdots M_m x| < |x| \tag{6.68}$$

In view of (6.64) and the assumption that $x \neq 0$, there must be a largest integer $k \in \{1, 2, \ldots, m\}$ such that $x \notin$ kernel $(M_k - I)$. We claim that

$$|M_k x| < |x| \tag{6.69}$$

To show that this is so we exploit the symmetry of M_k to write x as $x = \alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_n y_n$ where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are real numbers and $\{y_1, y_2, \ldots, y_n\}$ is an orthonormal set of eigenvectors of M_k with real eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_n$. Note that $|\lambda_i| \leq 1,\ i \in \{1, 2, \ldots, n\}$, because $|M_k| \leq 1$. Next observe that since $M_k x = \alpha_1 \lambda_1 y_1 + \alpha_2 \lambda_2 y_2 + \cdots + \alpha_n \lambda_n y_n$ and $M_k x \neq x$, there must be at least one integer j such that $\alpha_j \lambda_j \neq \alpha_j$. Hence $|\alpha_j \lambda_j y_j| < |\alpha_j y_j|$. But $|M_k x|^2 = |\alpha_1 \lambda_1 y_1|^2 + \cdots + |\alpha_j \lambda_j y_j|^2 + \cdots + |\alpha_n \lambda_n y_n|^2$ so

$$|M_k x|^2 < |\alpha_1 \lambda_1 y_1|^2 + \cdots + |\alpha_j y_j|^2 + \cdots + |\alpha_n \lambda_n y_n|^2$$

Moreover

$$|\alpha_1 \lambda_1 y_1|^2 + \cdots + |\alpha_j y_j|^2 + \cdots + |\alpha_n \lambda_n y_n|^2 \leq |\alpha_1 y_1|^2 + \cdots + |\alpha_j y_j|^2 + \cdots + |\alpha_n y_n|^2 = |x|^2$$

so $|M_k x|^2 < |x|^2$; therefore (6.69) is true.

In view of the definition of k, $M_j x = x, j \in \{k+1, \ldots, m\}$. From this and (6.69) it follows that $|M_1 \cdots M_m x| = |M_1 \cdots M_k x| \leq |M_1 \cdots M_{k-1}||M_k x| < |M_1 \cdots M_{k-1}||x|$. But $|M_1 \cdots M_{k-1}| \leq 1$ because each M_i has an induced 2 norm not exceeding 1. Therefore (6.68) is true. ■

6.3 Measurement Delays

In this section we consider a modified version of the flocking problem in which integer valued delays occur in sensing the values of headings which are available to agents. More precisely we suppose that at each time $t \in \{0, 1, 2, \ldots\}$, the value of neighboring agent j's headings which agent i may sense is $\theta_j(t - d_{ij}(t))$ where $d_{ij}(t)$ is a delay whose value at t is some integer between 0 and $m_j - 1$; here m_j is a pres-specified positive integer. While well established principles of feedback control would suggest that delays should be dealt with using dynamic compensation, in these notes we will consider the situation in which the delayed value of agent j's heading sensed by agent i at time t is the value which will be used in the heading update law for agent i. Thus

$$\theta_i(t+1) = \frac{1}{n_i(t)} \left(\sum_{j \in \mathcal{N}_i(t)} \theta_j(t - d_{ij}(t)) \right) \tag{6.70}$$

where $d_{ij}(t) \in \{0, 1, \ldots, (m_j - 1)\}$ if $j \neq i$ and $d_{ij}(t) = 0$ if $i = j$.

It is possible to represent this agent system using a state space model similar to the model discussed earlier for the delay-free case. Towards this end, let $\bar{\mathcal{D}}$ denote the set of all directed graphs with vertex set $\bar{\mathcal{V}} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \cdots \cup \mathcal{V}_n$ where $\mathcal{V}_i = \{v_{i1} \ldots, v_{im_i}\}$. Here vertex v_{i1} represents agent i and $\mathcal{V}_i$ is the

set of vertices associated with agent i. We sometimes write i for v_{i1}, $i \in \{1,2,\ldots,n\}$, and $\mathcal{V}$ for the subset of agent vertices $\{v_{11}, v_{21}, \ldots, v_{n1}\}$. Let $\bar{\mathcal{Q}}$ be an index set parameterizing $\bar{\mathcal{D}}$; i.e, $\bar{\mathcal{D}} = \{\mathbb{G}_q : q \in \bar{\mathcal{Q}}\}$

To represent the fact that each agent can use its own current heading in its update formula (6.70), we will utilize those graphs in $\bar{\mathcal{D}}$ which have self arcs at each vertex in $\mathcal{V}$. We will also require the arc set of each such graph to have, for $i \in \{1,2,\ldots,n\}$, an arc from each vertex $v_{ij} \in \mathcal{V}_i$ except the last, to its successor $v_{i(j+1)} \in \mathcal{V}_i$. Finally we stipulate that for each $i \in \{1,2,\ldots,n\}$, each vertex v_{ij} with $j > 1$ has in-degree of exactly 1. In the sequel we write $\mathcal{D}$ for the subset of all such graphs. Thus unlike the class of graphs $\mathcal{G}$ considered before, there are graphs in $\mathcal{D}$ possessing vertices without self-arcs. Nonetheless each vertex of each graph in $\mathcal{D}$ has positive in-degree. In the sequel we use the symbol $\mathcal{Q}$ to denote that subset of $\bar{\mathcal{Q}}$ for which $\mathcal{D} = \{\mathbb{G}_q : q \in \mathcal{Q}\}$.

The specific graph representing the sensed headings the agents use at time t to update their own headings according to (6.70), is that graph $\mathbb{G}_q \in \mathcal{D}$ whose arc set contains an arc from $v_{ik} \in \mathcal{V}_i$ to $j \in \mathcal{V}$ if agent j uses $\theta_i(t+1-k)$ to update. The set of agent heading update rules defined by (6.70) can now be written in state form. Towards this end define $\theta(t)$ to be that $(m_1 + m_2 + \cdots + m_i)$ vector whose first m_1 elements are $\theta_1(t)$ to $\theta_1(t+1-m_1)$, whose next m_2 elements are $\theta_2(t)$ to $\theta_2(t+1-m_2)$ and so on. Order the vertices of $\bar{\mathcal{V}}$ as $v_{11}, \ldots, v_{1m_1}, v_{21}, \ldots, v_{2m_2}, \ldots, v_{n1}, \ldots, v_{nm_n}$ and with respect to this ordering define

$$F_q = D_q^{-1} A_q', \quad q \in \mathcal{Q} \tag{6.71}$$

where A_q' is the transpose of the adjacency matrix the of $\mathbb{G}_q \in \mathcal{D}$ and D_q the diagonal matrix whose ijth diagonal element is the in-degree of vertex v_{ij} within the graph. Then

$$\theta(t+1) = F_{\sigma(t)}\theta(t), \quad t \in \{0,1,2,\ldots\} \tag{6.72}$$

where $\sigma : \{0,1,\ldots\} \to \mathcal{Q}$ is a switching signal whose value at time t, is the index of the graph representing which headings the agents use at time t to update their own headings according to (6.70). As before our goal is to characterize switching signals for which all entries of $\theta(t)$ converge to a common steady state value.

There are a number of similarities and a number of differences between the situation under consideration here and the delay-free situation considered earlier. For example, the notion of graph composition defined earlier can be defined in the obvious way for graphs in $\bar{\mathcal{D}}$. On the other hand, unlike the situation in the delay-free case, the set of graphs used to model the system under consideration, namely $\mathcal{D}$, is not closed under composition except in the special case when all of the delays are at most 1; i.e., when all of the $m_i \leq 2$. In order to characterize the smallest subset of $\bar{\mathcal{D}}$ containing $\mathcal{D}$ which is closed under composition, we will need several new concepts.

Hierarchical Graphs

As before, let $\bar{\mathcal{G}}$ be the set of all directed graphs with vertex set $\mathcal{V} = \{1, 2, \ldots n\}$. Let us agree to say that a rooted graph $\mathbb{G} \in \bar{\mathcal{G}}$ is a *hierarchical graph* with *hierarchy* $\{v_1, v_2, \ldots, v_n\}$ if it is possible to re-label the vertices in $\mathcal{V}$ as $v_1, v_2, \ldots v_n$ in such a way so that v_1 is a root of $\mathbb{G}$ with a self-arc and for $i > 1$, v_i has a neighbor v_j "lower " in the hierarchy where by *lower* we mean $j < i$. It is clear that any graph in $\bar{\mathcal{G}}$ with a root possessing a self-arc is hierarchial. Note that a graph may have more than one hierarchy and two graphs with the same hierarchy need not be equal. Note also that even though rooted graphs with the same hierarchy share a common root, examples show that the composition of hierarchial graphs in $\bar{\mathcal{G}}$ need not be hierarchial or even rooted. On the other hand the composition of two rooted graphs in $\bar{\mathcal{G}}$ with the same hierarchy is always a graph with the same hierarchy. To understand why this is so, consider two graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ in $\bar{\mathcal{G}}$ with the same hierarchy $\{v_1, v_2, \ldots, v_n\}$. Note first that v_1 has a self -arc in $\mathbb{G}_2 \circ \mathbb{G}_1$ because v_1 has self arcs in $\mathbb{G}_1$ and $\mathbb{G}_2$. Next pick any vertex v_i in $\mathcal{V}$ other than v_1. By definition, there must exist vertex v_j lower in the hierarchy than v_i such that (v_j, v_i) is an arc of $\mathbb{G}_2$. If $v_j = v_1$, then (v_1, v_i) is an arc in $\mathbb{G}_2 \circ \mathbb{G}_1$ because v_1 has a self-arc in $\mathbb{G}_1$. On the other hand, if $v_j \neq v_1$, then there must exist a vertex v_k lower in the hierarchy than v_j such that (v_k, v_j) is an arc of $\mathbb{G}_1$. It follows from the definition of composition that in this case (v_k, v_i) is an arc in $\mathbb{G}_2 \circ \mathbb{G}_1$. Thus v_i has a neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$ which is lower in the hierarchy than v_i. Since this is true for all v_i, $\mathbb{G}_2 \circ \mathbb{G}_1$ must have the same hierarchy as $\mathbb{G}_1$ and $\mathbb{G}_2$. This proves the claim that composition of two rooted graphs with the same hierarchy is a graph with the same hierarchy.

Our objective is to show that the composition of a sufficiently large number of graphs in $\bar{\mathcal{G}}$ with the same hierarchy is strongly rooted. Note that Proposition 6.3 cannot be used to reach this conclusion, because the v_i in the graphs under consideration here do not all necessarily have self-arcs.

As before, let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two graphs in $\bar{\mathcal{G}}$ with the same hierarchy $\{v_1, v_2, \ldots,$
$v_n\}$. Let v_i be any vertex in the hierarchy and suppose that v_j is a neighbor vertex of v_i in $\mathbb{G}_2$. If $v_j = v_1$, then v_i retains v_1 as a neighbor in the composition $\mathbb{G}_2 \circ \mathbb{G}_1$ because v_1 has a self-arc in $\mathbb{G}_1$. On the other hand, if $v_j \neq v_1$, then v_j has a neighboring vertex v_k in $\mathbb{G}_1$ which is lower in the hierarchy than v_j. Since v_k is a neighbor of v_i in the composition $\mathbb{G}_2 \circ \mathbb{G}_1$, we see that in this case v_i has acquired a neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$ lower in the hierarchy than a neighbor it had in $\mathbb{G}_2$. In summary, any vertex $v_i \in \mathcal{V}$ either has v_1 as neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$ or has a neighbor in $\mathbb{G}_2 \circ \mathbb{G}_1$ which is one vertex lower in the hierarchy than any neighbor it had in $\mathbb{G}_2$.

Now consider three graphs $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3$ in $\bar{\mathcal{G}}$ with the same hierarchy. By the same reasoning as above, any vertex $v_i \in \mathcal{V}$ either has v_1 as neighbor in $\mathbb{G}_3 \circ \mathbb{G}_2 \circ \mathbb{G}_1$ or has a neighbor in $\mathbb{G}_3 \circ \mathbb{G}_2 \circ \mathbb{G}_1$ which is one vertex lower in the hierarchy than any neighbor it had in $\mathbb{G}_3 \circ \mathbb{G}_2$. Similarly v_i either has v_1 as

neighbor in $\mathbb{G}_3 \circ \mathbb{G}_2$ or has a neighbor in $\mathbb{G}_3 \circ \mathbb{G}_2$ which is one vertex lower in the hierarchy than any neighbor it had in $\mathbb{G}_3$. Combining these two observations we see that any vertex $v_i \in \mathcal{V}$ either has v_1 as neighbor in $\mathbb{G}_3 \circ \mathbb{G}_2 \circ \mathbb{G}_1$ or has a neighbor in $\mathbb{G}_3 \circ \mathbb{G}_2 \circ \mathbb{G}_1$ which is two vertices lower in the hierarchy than any neighbor it had in $\mathbb{G}_3$. This clearly generalizes and so after the composition of m such graphs $\mathbb{G}_1, \mathbb{G}_2, \ldots \mathbb{G}_m$, v_i either has v_1 as neighbor in $\mathbb{G}_m \circ \cdots \mathbb{G}_2 \circ \mathbb{G}_1$ or has a neighbor in $\mathbb{G}_m \circ \cdots \mathbb{G}_2 \circ \mathbb{G}_1$ which is $m-1$ vertices lower in the hierarchy than any neighbor it had in $\mathbb{G}_m$. It follows that if $m \geq n$, then v_i must be a neighbor of v_1. Since this is true for all vertices, we have proved the following.

Proposition 6.14. *Let $\mathbb{G}_1, \mathbb{G}_2, \ldots \mathbb{G}_m$ denote a set of rooted graphs in $\bar{\mathcal{G}}$ which all have the same hierarchy. If $m \geq n-1$ then $\mathbb{G}_m \circ \cdots \mathbb{G}_2 \circ \mathbb{G}_1$ is strongly rooted.*

As we've already pointed out, Proposition 6.14 is not a consequence of Proposition 6.3 because Proposition 6.3 requires all vertices of all graphs in the composition to have self-arcs whereas Proposition 6.14 does not. On the other hand, Proposition 6.3 is not a consequence of Proposition 6.14 because Proposition 6.14 only applies to graphs with the same hierarchy whereas Proposition 6.3 does not.

Delay Graphs

We now return to the study of the graphs in $\mathcal{D}$. As before $\mathcal{D}$ is the subset of $\bar{\mathcal{D}}$ consisting of those graphs which (i) have self arcs at each vertex in $\mathcal{V} = \{1, 2, \ldots, \}$, (ii) for each $i \in \{1, 2, \ldots, n\}$, have an arc from each vertex $v_{ij} \in \mathcal{V}_i$ except the last, to its successor $v_{i(j+1)} \in \mathcal{V}_i$, and (iii) for each $i \in \{1, 2, \ldots, n\}$, each vertex v_{ij} with $j > 1$ has in-degree of exactly 1. It can easily be shown by example that $\mathcal{D}$ is not closed under composition. We deal with this problem as follows. A graph $\mathbb{G} \in \bar{\mathcal{D}}$ is said to be a *delay graph* if for each $i \in \{1, 2, \ldots, n\}$, (i) every neighbor of $\mathcal{V}_i$ which is not in $\mathcal{V}_i$ is a neighbor of v_{i1} and (ii) the subgraph of $\mathbb{G}$ induced by $\mathcal{V}_i$ has $\{v_{i1} \ldots, v_{im_i}\}$ as a hierarchy. It is easy to see that every graph in $\mathcal{D}$ is a delay graph. More is true.

Proposition 6.15. *The set of delay graphs in $\bar{\mathcal{D}}$ is closed under composition.*

To prove this proposition, we will need the following fact.

Lemma 6.18. *Let $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_q$ be any sequence of $q > 1$ directed graphs in $\bar{\mathcal{G}}$. For $i \in \{1, 2, \ldots, q\}$, let $\bar{\mathbb{G}}_i$ be the subgraph of $\mathbb{G}_i$ induced by $\mathcal{S} \subset \mathcal{V}$. Then $\bar{\mathbb{G}}_q \circ \cdots \circ \bar{\mathbb{G}}_2 \circ \bar{\mathbb{G}}_1$ is contained in the subgraph of $\mathbb{G}_q \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1$ induced by $\mathcal{S}$.*

Proof of Lemma 6.18: It will be enough to prove the lemma for $q = 2$, since the proof for $q > 2$ would then directly follow by induction. Suppose $q = 2$.

Let (i,j) be in $\mathcal{A}(\bar{\mathbb{G}}_2 \circ \bar{\mathbb{G}}_1)$. Then $i,j \in \mathcal{S}$ and there exists an integer $k \in \mathcal{S}$ such that $(i,k) \in \mathcal{A}(\bar{\mathbb{G}}_1)$ and $(k,j) \in \mathcal{A}(\bar{\mathbb{G}}_2)$. Therefore $(i,k) \in \mathcal{A}(\mathbb{G}_1)$ and $(k,j) \in \mathcal{A}(\mathbb{G}_2)$. Thus $(i,j) \in \mathcal{A}(\mathbb{G}_2 \circ \mathbb{G}_1)$. But $i,j \in \mathcal{S}$ so (i,j) must be an arc in the subgraph of $\mathbb{G}_2 \circ \mathbb{G}_1$ induced by $\mathcal{S}$. Since this clearly is true for all arcs in $\mathcal{A}(\bar{\mathbb{G}}_2 \circ \bar{\mathbb{G}}_1)$, the proof is complete. ■

Proof of Proposition 6.15: Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two delay graphs in $\bar{\mathcal{D}}$. It will first be shown that for each $i \in \{1,2,\ldots,n\}$, every neighbor of $\mathcal{V}_i$ which is not in $\mathcal{V}_i$ is a neighbor of v_{i1} in $\mathbb{G}_2 \circ \mathbb{G}_1$. Fix $i \in \{1,2,\ldots,n\}$ and let v be a neighbor of $\mathcal{V}_i$ in $\mathbb{G}_2 \circ \mathbb{G}_1$ which is not in $\mathcal{V}_i$. Then $(v,k) \in \mathcal{A}(\mathbb{G}_2 \circ \mathbb{G}_1)$ for some $k \in \mathcal{V}_i$. Thus there is a $s \in \bar{\mathcal{V}}$ such that $(v,s) \in \mathcal{A}(\mathbb{G}_1)$ and $(s,k) \in \mathcal{A}(\mathbb{G}_2)$. If $s \notin \mathcal{V}_i$, then $(s,v_{i1}) \in \mathcal{A}(\mathbb{G}_2)$ because $\mathbb{G}_2$ is a delay graph. Thus in this case $(v,v_{i1}) \in \mathcal{A}(\mathbb{G}_2 \circ \mathbb{G}_1)$ because of the definition of composition. If, on the other hand, $s \in \mathcal{V}_i$, then $(v,v_{i1}) \in \mathcal{A}(\mathbb{G}_1)$ because $\mathbb{G}_1$ is a delay graph. Thus in this case $(v,v_{i1}) \in \mathcal{A}(\mathbb{G}_2 \circ \mathbb{G}_1)$ because v_{i1} has a self-arc in $\mathbb{G}_2$. This proves that every neighbor of $\mathcal{V}_i$ which is not in $\mathcal{V}_i$ is a neighbor of v_{i1} in $\mathbb{G}_2 \circ \mathbb{G}_1$. Since this must be true for each $i \in \{1,2,\ldots,n\}$, $\mathbb{G}_2 \circ \mathbb{G}_1$ has the first property defining delay graphs in $\bar{\mathcal{D}}$.

To establish the second property, we exploit the fact that the composition of two graphs with the same hierarchy is a graph with the same hierarchy. Thus for any integer $i \in \{1,2,\ldots,n\}$, the composition of the subgraphs of $\mathbb{G}_1$ and $\mathbb{G}_2$ respectively induced by $\mathcal{V}_i$ must have the hierarchy $\{v_{i1},v_{i2},\ldots,v_{im_i}\}$. But by Lemma 6.18, for any integer $i \in \{1,2,\ldots,n\}$, the composition of the subgraphs of $\mathbb{G}_1$ and $\mathbb{G}_2$ respectively induced by $\mathcal{V}_i$, is contained in the subgraph of the composition of $\mathbb{G}_1$ and $\mathbb{G}_2$ induced by $\mathcal{V}_i$. This implies that for $i \in \{1,2,\ldots,n\}$, the subgraph of the composition of $\mathbb{G}_1$ and $\mathbb{G}_2$ induced by $\mathcal{V}_i$ has $\{v_{i1},v_{i2},\ldots,v_{im_i}\}$ as a hierarchy. ■

In the sequel we will state and prove conditions under which the composition of a sequence of delay graphs is strongly rooted. To do this, we will need to introduce several concepts. By the *quotient graph* of $\mathbb{G} \in \bar{\mathcal{D}}$, is meant that directed graph with vertex set $\mathcal{V}$ whose arc set consists of those arcs (i,j) for which $\mathbb{G}$ has an arc from some vertex in $\mathcal{V}_i$ to some vertex in $\mathcal{V}_j$. The quotient graph of $\mathbb{G}$ models which headings are being used by each agent in updates without describing the specific delayed headings actually being used. Our main result regarding delay graphs is as follows.

Proposition 6.16. *Let m be the largest integer in the set $\{m_1,m_2,\ldots,m_n\}$. The composition of any set of at least $m(n-1)^2+m-1$ delay graphs will be strongly rooted if the quotient graph of each of the graphs in the composition is rooted.*

To prove this proposition we will need several more concepts. Let us agree to say that a delay graph $\mathbb{G} \in \bar{\mathcal{D}}$ has *strongly rooted hierarchies* if for each $i \in \mathcal{V}$, the subgraph of $\mathbb{G}$ induced by $\mathcal{V}_i$ is strongly rooted. Proposition 6.14 states that a hierarchial graph on m_i vertices will be strongly rooted if it is the composition of at least $m_i - 1$ rooted graphs with the same hierarchy. This and Lemma 6.18 imply that the subgraph of the composition of at least

$m_i - 1$ delay graphs induced by $\mathcal{V}_i$ will be strongly rooted. We are led to the following lemma.

Lemma 6.19. *Any composition of at least $m-1$ delay graphs in $\bar{\mathcal{D}}$ has strongly rooted hierarchies.*

To proceed we will need one more type of graph which is uniquely determined by a given graph in $\bar{\mathcal{D}}$. By the *agent subgraph* of $\mathbb{G} \in \bar{\mathcal{D}}$ is meant the subgraph of $\mathbb{G}$ induced by $\mathcal{V}$. Note that while the quotient graph of $\mathbb{G}$ describes relations between distinct agent hierarchies, the agent subgraph of $\mathbb{G}$ only captures the relationships between the roots of the hierarchies.

Lemma 6.20. *Let $\mathbb{G}_p$ and $\mathbb{G}_q$ be delay graphs in $\bar{\mathcal{D}}$. If $\mathbb{G}_p$ has a strongly rooted agent subgraph and $\mathbb{G}_q$ has strongly rooted hierarchies, then the composition $\mathbb{G}_q \circ \mathbb{G}_p$ is strongly rooted.*

Proof of Lemma 6.20: Let v_{i1} be a root of the agent subgraph of $\mathbb{G}_p$ and let v_{jk} be any vertex in $\bar{\mathcal{V}}$. Then $(v_{i1}, v_{j1}) \in \mathcal{A}(\mathbb{G}_p)$ because the agent subgraph of $\mathbb{G}_p$ is strongly rooted. Moreover, $(v_{j1}, v_{jk}) \in \mathcal{A}(\mathbb{G}_q)$ because $\mathbb{G}_q$ has strongly rooted hierarchies. Therefore, in view of the definition of graph composition $(v_{i1}, v_{jk}) \in \mathcal{A}(\mathbb{G}_q \circ \mathbb{G}_p)$. Since this must be true for every vertex in $\bar{\mathcal{V}}$, $\mathbb{G}_q \circ \mathbb{G}_p$ is strongly rooted. ■

Lemma 6.21. *The agent subgraph of any composition of at least $(n-1)^2$ delay graphs in $\bar{\mathcal{D}}$ will be strongly rooted if the agent subgraph of each of the graphs in the composition is rooted.*

Proof of Lemma 6.21: Let $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_q$ be any sequence of $q \geq (n-1)^2$ delay graphs in $\bar{\mathcal{D}}$ whose agent subgraphs, $\bar{\mathbb{G}}_i \;\; i \in \{1, 2, \ldots, q\}$, are all rooted. By Proposition 6.10, $\bar{\mathbb{G}}_q \circ \cdots \circ \bar{\mathbb{G}}_2 \circ \bar{\mathbb{G}}_1$ is strongly rooted. But $\bar{\mathbb{G}}_q \circ \cdots \circ \bar{\mathbb{G}}_2 \circ \bar{\mathbb{G}}_1$ is contained in the agent subgraph of $\mathbb{G}_q \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1$ because of Lemma 6.18. Therefore the agent subgraph of $\mathbb{G}_q \circ \cdots \circ \mathbb{G}_2 \circ \mathbb{G}_1$ is strongly rooted. ■

Lemma 6.22. *Let $\mathbb{G}_p$ and $\mathbb{G}_q$ be delay graphs in $\bar{\mathcal{D}}$. If $\mathbb{G}_p$ has a strongly rooted hierarchies and $\mathbb{G}_q$ has a rooted quotient graph, then the agent subgraph of the composition $\mathbb{G}_q \circ \mathbb{G}_p$ is rooted.*

Proof of Lemma 6.22: Let (i, j) be any arc in the quotient graph of $\mathbb{G}_q$ with $i \neq j$. This means that $(v_{ik}, v_{js}) \in \mathcal{A}(\mathbb{G}_q)$ for some $v_{ik} \in \mathcal{V}_i$ and $v_{js} \in \mathcal{V}_j$. Clearly $(v_{i1}, v_{ik}) \in \mathcal{A}(\mathbb{G}_p)$ because $\mathbb{G}_p$ has strongly rooted hierarchies. Moreover since $i \neq j$, v_{ik} is a neighbor of $\mathcal{V}_j$ which is not in $\mathcal{V}_j$. From this and the definition of a delay graph, it follows that v_{ik} is a neighbor of v_{j1}. Therefore $(v_{ik}, v_{j1}) \in \mathcal{A}(\mathbb{G}_q)$. Thus $(v_{i1}, v_{j1}) \in \mathcal{A}(\mathbb{G}_q \circ \mathbb{G}_p)$. We have therefore proved that for any path of length one between any two distinct vertices i, j in the quotient graph of $\mathbb{G}_q$, there is a corresponding path between vertices v_{i1} and v_{j1} in the agent subgraph of $\mathbb{G}_q \circ \mathbb{G}_p$. This implies that for any path of any length between any two distinct vertices i, j in the quotient graph of

$\mathbb{G}_q$, there is a corresponding path between vertices v_{i1} and v_{j1} in the agent subgraph of $\mathbb{G}_q \circ \mathbb{G}_p$. Since by assumption, the quotient graph of $\mathbb{G}_q$ is rooted, the agent subgraph of $\mathbb{G}_q \circ \mathbb{G}_p$ must be rooted as well. ■

Proof of Proposition 6.16: Let $\mathbb{G}_1, \mathbb{G}_2, \ldots \mathbb{G}_s$ be a sequence of at least $m(n-1)^2+m-1$ delay graphs with strongly rooted quotient graphs. The graph $\mathbb{G}_s \circ \cdots \mathbb{G}_{(m(n-1)^2+1)}$ is composed of at least $m-1$ delay graphs. Therefore $\mathbb{G}_s \circ \cdots \mathbb{G}_{(m(n-1)^2+1)}$ must have strongly rooted hierarchies because of Lemma 6.19. In view of Lemma 6.20, to complete the proof it is enough to show that $\mathbb{G}_{(m(n-1)^2} \circ \cdots \circ \mathbb{G}_1$ has a strongly rooted agent subgraph. But $\mathbb{G}_{(m(n-1)^2} \circ \cdots \circ \mathbb{G}_1$ is the composition of $(n-1)^2$ graphs, each itself a composition of m delay graphs with rooted quotient graphs. In view of Lemma 6.21, to complete the proof it is enough to show that the agent subgraph of any composition of m delay graphs is rooted if each of the quotient graph of each delay graph in the composition is rooted. Let $\mathbb{H}_1, \mathbb{H}_2, \ldots, \mathbb{H}_m$ be such a family of delay graphs. By assumption, $\mathbb{H}_m$ has a rooted quotient graph. In view of Lemma 6.22, the agent subgraph of $\mathbb{H}_m \circ \mathbb{H}_{m-1} \circ \cdots \circ \mathbb{H}_1$ will be rooted if $\mathbb{H}_{m-1} \circ \cdots \circ \mathbb{H}_1$ has strongly rooted hierarchies. But $\mathbb{H}_{m-1} \circ \cdots \circ \mathbb{H}_1$ has this property because of Lemma 6.19. ■

Convergence

Using the results from the previous section, it is possible to state results for the flocking problem with measurement delays similar to those discussed earlier for the delay free case. Towards this end let us agree to say that a finite sequence of graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ in $\mathcal{D}$ is *jointly quotient rooted* if the quotient of the composition $\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{(k-1)}} \circ \cdots \circ \mathbb{G}_{p_1}$ is rooted.

In the sequel we will say that an infinite sequence of graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}, \ldots,$ in $\mathcal{D}$ is *repeatedly jointly quotient rooted* if there is a positive integer m for which each finite sequence $\mathbb{G}_{p_{m(k-1)+1}}, \ldots, \mathbb{G}_{p_{mk}}, \quad k \geq 1$ is jointly quotient rooted. We are now in a position to state our main result on leaderless coordination with measurement delays.

Theorem 6.5. *Let $\theta(0)$ be fixed and with respect to (6.72), let $\sigma : [0, 1, 2, \ldots) \to \bar{\mathcal{Q}}$ be a switching signal for which the infinite sequence of graphs $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ in $\mathcal{D}$ is repeatedly jointly rooted. Then there is a constant steady state heading θ_{ss}, depending only on $\theta(0)$ and σ, for which*

$$\lim_{t \to \infty} \theta(t) = \theta_{ss} \mathbf{1} \tag{6.73}$$

where the limit is approached exponentially fast.

The proof of this theorem exploits Proposition 6.16 and parallels exactly the proof of Theorem 6.3. A proof of Theorem 6.5 therefore will not be given.

6.4 Asynchronous Flocking

In this section we consider a modified version of the flocking problem in which each agent independently updates its heading at times determined by its own clock [44]. We do not assume that the groups' clocks are synchronized together or that the times any one agent updates its heading are evenly spaced. Updating of agent i's heading is done as follows. At its *kth sensing event time* t_{ik}, agent i senses the headings $\theta_j(t_{ik}),\ \ j \in \mathcal{N}_i(t_{ik})$ of its current neighbors and from this data computes its *kth* way-point $w_i(t_{ik})$. In the sequel we will consider way point rules based on averaging. In particular

$$w_i(t_{ik}) = \frac{1}{n_i(t_{ik})} \left(\sum_{i \in \mathcal{N}_i(t_{ik})} \theta_j(t_{ik}) \right), \quad i\{1, 2, \ldots, n\} \tag{6.74}$$

where $n_i(t_{ik})$ is the number of neighbor of elements in neighbor index set $\mathcal{N}_i(t_{ik})$. Agent i then changes its heading from $\theta_i(t_{ik})$ to $w_i(t_{ik})$ on the interval $[t_{ik}, t_{i(k+1)})$. In these notes we will consider the case each agent updates its headings instantaneously at its own even times, and that it maintains fixed headings between its event times. More precisely, we will assume that agent i reaches its kth way-point at its $(k+1)$st event time and that $\theta_i(t)$ is constant on each continuous-time interval $(t_{i(k-1)}, t_{ik}]$, $k \geq 1$, where $t_{i0} = 0$ is agent i's zeroth event time. In other words for $k \geq 0$, agent i's heading satisfies is

$$\theta_i(t_{i(k+1)}) = \frac{1}{n_i(t_{ik})} \left(\sum_{j \in \mathcal{N}_i(t_{ik})} \theta_j(t_{ik}) \right) \tag{6.75}$$

$$\theta_i(t) = \theta_i(t_{ik}), \quad t_{i(k-1)} < t \leq t_{ik}, \tag{6.76}$$

To ensure that each agent's neighbors are unambiguously defined at each of its event times, we will further assume that agents move continuously.

Analytic Synchronization

To develop conditions under which all agents eventually move with the same heading requires the analysis of the asymptotic behavior of the *asynchronous* process which the $2n$ heading equations of the form (6.75), (6.76) define. Despite the apparent complexity of this process, it is possible to capture its salient features using a suitably defined *synchronous* discrete-time, hybrid dynamical system $\mathbb{S}$. We call the sequence of steps involved in defining $\mathbb{S}$ *analytic synchronization*. Analytic synchronization is applicable to any finite family of continuous or discrete time dynamical processes $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \ldots, \mathbb{P}_n\}$ under the following conditions. First, each process $\mathbb{P}_i$ must be a dynamical system whose inputs consist of functions of the states of the other processes as well as signals which are exogenous to the entire family. Second, each

process $\mathbb{P}_i$ must have associated with it an ordered sequence of event times $\{t_{i1}, t_{i2}, \ldots\}$ defined in such a way so that the state of $\mathbb{P}_i$ at event time $t_{i(k_i+1)}$ is uniquely determined by values of the exogenous signals and states of the $\mathbb{P}_j, \quad j \in \{1, 2, \ldots, n\}$ at event times t_{jk_j} which occur prior to $t_{i(k_i+1)}$ but in the finite past. Event time sequences for different processes need not be synchronized. Analytic synchronization is a procedure for creating a single synchronous process for purposes of analysis which captures the salient features of the original n asynchronously functioning processes. As a first step, all n event time sequences are merged into a single ordered sequence of even times $\mathcal{T}$. The "synchronized" state of $\mathbb{P}_i$ is then defined to be the original of $\mathbb{P}_i$ at $\mathbb{P}_i$'s event times $\{t_{i1}, t_{i2}, \ldots\}$ plus possibly some additional variables; at values of $t \in \mathcal{T}$ between event times t_{ik_i} and $t_{i(k_i+1)}$, the synchronized state of $\mathbb{P}_i$ is taken to be the same at the value of its original state at time $t_{i(k+1)}$. Although it is not always possible to carry out all of these steps, when it is what ultimately results is a synchronous dynamical system $\mathbb{S}$ evolving on the index set of $\mathcal{T}$, with state composed of the synchronized states of the n individual processes under consideration. We now use these ideas to develop such a synchronous system $\mathbb{S}$ for the asynchronous process we've been studying.

Definition of $\mathbb{S}$

As a first step, let $\mathcal{T}$ denote the set of all event times of all n agents. Relabel the elements of $\mathcal{T}$ as $t_0, t_1, t_2, \cdots$ in such a way so that $t_j < t_{j+1}, \ j \in \{1, 2, \ldots\}$. Next define

$$\bar{\theta}_i(\tau) = \theta_i(t_\tau), \quad \tau \geq 0, \quad i \in \{1, 2, \ldots, n\} \tag{6.77}$$

In view of (6.75), it must be true that if t_τ is an event time of agent i, then

$$\bar{\theta}_i(\tau') = \frac{1}{\bar{n}_i(t_\tau)} \left(\sum_{j \in \bar{\mathcal{N}}_i(\tau)} \bar{\theta}_j(\tau) \right)$$

where $\bar{\mathcal{N}}_i(\tau) = \mathcal{N}_i(t_\tau)$, $\bar{n}_i(\tau) = n_i(t_\tau)$ and $t_{\tau'}$ is the next event time of agent i after t_τ. But $\bar{\theta}_i(\tau') = \bar{\theta}_i(\tau + 1)$ because $\theta_i(t)$ is constant for $t_\tau < t \leq t_{\tau'}$ {cf., (6.76)}. Therefore

$$\bar{\theta}_i(\tau + 1) = \frac{1}{\bar{n}_i(t_\tau)} \left(\sum_{j \in \bar{\mathcal{N}}_i(\tau)} \bar{\theta}_j(\tau) \right) \tag{6.78}$$

if t_τ is an event time of agent i. Meanwhile if t_τ is not an event time of agent i, then

$$\bar{\theta}_i(\tau + 1) = \bar{\theta}_i(\tau), \tag{6.79}$$

again because $\theta_i(t)$ is constant between event times. Note that if we *define* $\bar{\mathcal{N}}_i(\tau) = \{i\}$ and $\bar{n}_i(\tau) = 1$ for every value of τ for which t_τ is not an event time of agent i, then (6.79) can be written as

$$\bar{\theta}_i(\tau+1) = \frac{1}{\bar{n}_i(t_\tau)} \left(\sum_{j \in \bar{\mathcal{N}}_i(\tau)} \bar{\theta}_j(\tau) \right) \tag{6.80}$$

Doing this enables us to combine (6.78) and (6.80) into a single formula valid for all $\tau \geq 0$. In other words, agent i's heading satisfies

$$\bar{\theta}_i(\tau+1) = \frac{1}{\bar{n}_i(t_\tau)} \left(\sum_{j \in \bar{\mathcal{N}}_i(\tau)} \bar{\theta}_j(\tau) \right), \quad \tau \geq 0 \tag{6.81}$$

where

$$\bar{\mathcal{N}}_i(\tau) = \left\{ \begin{array}{l} \mathcal{N}_i(t_\tau) \text{ if } t_\tau \text{ is an event time of agent } i \\ \\ \{i\} \quad \text{if } t_\tau \text{ is not an event time of agent } i \end{array} \right\} \tag{6.82}$$

and $\bar{n}_i(\tau)$ is the number of indices in $\bar{\mathcal{N}}_i(\tau)$. For purposes of analysis, it is useful to interpret (6.82) as meaning that between agent i's event times, its only neighbor is itself. There are n equations of the form in (6.81) and together they define a synchronous system $\mathbb{S}$ which models the evolutions of the n agents' headings at event times.

State Space Model

As before, we can represent the neighbor relationships associated with (6.82) using a directed graph $\mathbb{G}$ with vertex set $\mathcal{V} = \{1, 2, \ldots n\}$ and arc $\mathcal{A}(\mathbb{G}) \subset \mathcal{V} \times \mathcal{V}$ which is defined in such a way so that (i, j) is an arc from i to j just in case agent i is a neighbor of agent j. Thus as before, $\mathbb{G}$ is a directed graph on n vertices with at most one arc from any vertex to another and with exactly one self - arc at each vertex. We continue to write $\mathcal{G}$ for the set of all such graphs and we also continue to use the symbol $\mathcal{P}$ to denote a set indexing $\mathcal{G}$.

For each $p \in \mathcal{P}$, let $F_p = D_p^{-1} A_p'$, where A_p' is the transpose of the adjacency matrix the of graph $\mathbb{G}_p \in \mathcal{G}$ and D_p the diagonal matrix whose jth diagonal element is the in-degree of vertex j within the graph. The set of agent heading update rules defined by (6.82) can be written in state form as

$$\bar{\theta}(\tau+1) = F_{\sigma(\tau)} \bar{\theta}(\tau), \quad \tau \in \{0, 1, 2, \ldots\} \tag{6.83}$$

where $\bar{\theta}$ is the heading vector $\bar{\theta} = \left(\bar{\theta}_1 \; \bar{\theta}_2 \; \ldots \; \bar{\theta}_n\right)'$, and $\sigma : \{0, 1, \ldots\} \to \mathcal{P}$ is a switching signal whose value at time τ, is the index of the graph representing the agents' neighbor relationships at time τ.

Up to this point things are essentially the same as in the basic flocking problem treated in Section 6.1. But when one considers the type of graphs in $\mathcal{G}$ which are likely to be encountered along a given trajectory, things are quite different. Note for example, that the only vertices of $\mathbb{G}_{\sigma(\tau)}$ which can

have more than one incoming arc, are those of agents for whom τ is an event time. Thus in the most likely situation when distinct agents have only distinct event times, there will be at most one vertex in each graph $\mathbb{G}_{\sigma(\tau)}$ which has more than one incoming arc. It is this situation we want to explore further. Toward this end, let $\mathcal{G}^* \subset \mathcal{G}$ denote the subclass of all graphs which have at most one vertex with more than one incoming arc. Note that for $n > 2$, there is no rooted graph in $\mathcal{G}^*$. Nonetheless, in the light of Theorem 6.3 it is clear that convergence to a common steady state heading will occur if the infinite sequence of graphs $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ is repeatedly jointly rooted. This of course would require that there exist a jointly rooted sequence of graphs from $\mathcal{G}^*$. We will now explain why such sequences do in fact exist.

Let us agree to call a graph $\mathbb{G} \in \mathcal{G}$ an *all neighbor graph centered at* v if every vertex of $\mathbb{G}$ is a neighbor of v. Thus $\mathbb{G}$ is an all neighbor graph centered at v if and only if its reverse $\mathbb{G}'$ is strongly rooted at v. Note that every all neighbor graph in $\mathcal{G}$ is also in $\mathcal{G}^*$. Note also that all neighbor graphs are maximal in $\mathcal{G}^*$ with respect to the partial ordering of $\mathcal{G}^*$ by inclusion. Note also the composition of any all neighbor graph with itself is itself. On the other hand, because of the union of two graphs in $\mathcal{G}$ is always contained in the composition of the two graphs, the composition of n all neighbor graphs with distinct centers must be a graph in which each vertex is a neighbor of every other; i.e., the complete graph. Thus the composition of n all neighbor graphs with distinct centers is strongly rooted. In summary, the hypothesis of Theorem 6.3 is not vacuous for the asynchronous problem under consideration. When that hypothesis is satisfied, convergence to a common steady state heading will occur.

6.5 Leader Following

In this section we consider two modified versions of the flocking problem for the same group n agents as before, but now with one of the group's members {say agent 1} acting as the group's *leader*. In the first version of the problem, the remaining agents, henceforth called *followers* and labelled 2 through n, do not know who the leader is or even if there is a leader. Accordingly they continue to use the same heading update rule (6.1) as before. The leader on the other hand, acting on its own, ignores update rule (6.1) and moves with a constant heading $\theta_1(0)$. Thus

$$\theta_1(t+1) = \theta_1(t) \tag{6.84}$$

The situation just described can be modelled as a state space system

$$\theta(t+1) = F_{\sigma(t)}\theta(t), \quad t \geq 0 \tag{6.85}$$

just as before, except now agent 1 is constrained to have no neighbors other than itself. The graphs $\mathbb{G}_p$ which model neighbor relations accordingly all have a distinguished *leader vertex* which has no incoming arcs other than its own.

Much like before, our goal here is to show for a large class of switching signals and for any initial set of follower agent headings, that the headings of all n followers converge to the heading of the leader. Convergence in the leaderless case under the most general, required the sequence of graphs $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ encountered along a trajectory to be repeatedly jointly rooted. For the leader follower case now under consideration, what's required is exactly the same. However, since the leader vertex has only one incoming arc, the only way $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ can be repeatedly jointly rooted, is that the sequence be "rooted at the leader vertex $v = 1$." More precisely, an infinite sequence of graphs $\mathbb{G}_{p_1}, \mathbb{G}_{p_2}$, in $\mathcal{G}$ is *repeatedly jointly rooted at* v if there is a positive integer m for which each finite sequence $\mathbb{G}_{p_{m(k-1)+1}}, \ \ldots, \mathbb{G}_{p_{mk}}, \quad k \geq 1$ is "jointly rooted at v"; a finite sequence of directed graphs $\mathbb{G}_{p_1}, \ \mathbb{G}_{p_2}, \ldots, \mathbb{G}_{p_k}$ is *jointly rooted at* v if the composition $\mathbb{G}_{p_k} \circ \mathbb{G}_{p_{k-1}} \circ \cdots \circ \mathbb{G}_{p_1}$ is rooted at v. Our main result on discrete-time leader following is next.

Theorem 6.6. *Let $\theta(0)$ be fixed and let $\sigma : [0, 1, 2, \ldots) \to \mathcal{P}$ be a switching signal for which the infinite sequence of graphs $\mathbb{G}_{\sigma(0)}, \mathbb{G}_{\sigma(1)}, \ldots$ is repeatedly jointly rooted. Then*

$$\lim_{t \to \infty} \theta(t) = \theta_1(0)\mathbf{1} \tag{6.86}$$

where the limit is approached exponentially fast.

Proof of Theorem 6.6: Since any sequence which is repeatedly jointly rooted at v is repeatedly jointly rooted, Theorem 6.3 is applicable. Therefore the headings of all n agents converge exponentially fast to a single common steady state heading θ_{ss}. But since the heading of the leader is fixed, $\theta_s s$ must be the leader's heading. ■

References

1. A. S. Morse. Control using logic-based switching. In A. Isidori, editor, *Trends in Control*, pages 69–113. Springer-Verlag, 1995.
2. Daniel Liberzon. *Switching in Systems and Control*. Birkhäuser, 2003.
3. V.D. Blondel and J.N. Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems and Control Letters*, 41:135–140, 2000.
4. J. P. Hespanha. Root-mean-square gains of switched linear systems. *IEEE Transactions on Automatic Control*, pages 2040–2044, Nov 2003.
5. J. P. Hespahna and A. S. Morse. Switching between stabilizing controllers. *Automatica*, 38(11), nov 2002.
6. I. M. Lie Ying. Adaptive disturbance rejection. *IEEE Transactions on Automatic Control*, 1991. Rejected for publication.
7. W. M. Wonham and A. S. Morse. Decoupling and pole assignment in linear multivariable systems: A geometric approach. *SIAM Journal on Control*, 8(1):1–18, February 1970.
8. M. A. Aizerman and F. R. Gantmacher. *Absolute Stability of Regulator Systems*. Holden-Day, 1964.

9. H. A. Simon. email communication, Feb. 23, 1997.
10. H. A. Simon. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica*, pages 171 –187, feb 1956.
11. J. P. Hespanha and A. S. Morse. Certainty equivalence implies detectability. *Systems and Control Letters*, pages 1–13, 1999.
12. J. P. Hespanha. *Logic - Based Switching Algorithms in Control*. PhD thesis, Yale University, 1998.
13. F. M. Pait and A. S. Morse. A cyclic switching strategy for parameter-adaptive control. *IEEE Transactions on Automatic Control*, 39(6):1172–1183, June 1994.
14. B. D. O. Anderson, T. S. Brinsmead, F. de Bruyne, J. P. Hespanha, D. Liberzon, and A. S. Morse. Multiple model adaptive control, part 1: Finite coverings. *International Journal on Robust and Nonlinear Control*, pages 909–929, September 2000.
15. F. M. Pait. On the topologies of spaces on linear dynamical systems commonly employed as models for adaptive and robust control design. In *Proceedings of the Third SIAM Conference on Control and Its Applications*, 1995.
16. G. Zames and A. K. El-Sakkary. Unstable systems and feedback: the gap metric. In *Proc. Allerton Conf.*, pages 380–385, 1980.
17. G. Vinnicombe. A ν-gap distance for uncertain and nonlinear systems. In *Proc. of the 38th Conf. on Decision and Contr.*, pages 2557–2562, 1999.
18. A. S. Morse. Supervisory control of families of linear set-point controllers - part 1: Exact matching. *IEEE Transactions on Automatic Control*, pages 1413–1431, oct 1996.
19. A. S. Morse. Supervisory control of families of linear set-point controllers - part 2: Robustness. *IEEE Transactions on Automatic Control*, 42:1500–1515, nov 1997. see also Proceedings of 1995 IEEE Conference on Decision and Control pp. 1750-1755.
20. S. R. Kulkarni and P. J. Ramadge. Model and controller selection policies based on output prediction errors. *IEEE Transactions on Automatic Control*, 41:1594–1604, 1996.
21. D. Borrelli, A. S. Morse, and E. Mosca. Discrete-time supervisory control of families of 2-degree of freedom linear set-point controllers. *IEEE Transactions on Automatic Control*, pages 178–181, jan 1999.
22. K. S. Narendra and J. Balakrishnan. Adaptive control using multiple models. *IEEE Transactions on Automatic Control*, pages 171 –187, feb 1997.
23. A. S. Morse. A bound for the disturbance-to-tracking error gain of a supervised set-point control system. In D Normand Cyrot, editor, *Perspectives in Control – Theory and Applications*, pages 23 – 41. Springer-Verlag, 1998.
24. A. S. Morse. Analysis of a supervised set-point control system containing a compact continuum of finite dimensional linear controllers. In *Proc. 2004 MTNS*, 2004.
25. T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, pages 1226–1229, 1995.
26. C. Reynolds. Flocks, birds, and schools: a distributed behavioral model. *Computer Graphics*, 21:25–34, 1987.
27. A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, pages 988–1001, june 2003. also in Proc. 2002 IEEE CDC, pages 2953 - 2958.

28. C. Godsil and G. Royle. *Algebraic Graph Theory*. Springer Graduate Texts in Mathematics # 207, New York, 2001.
29. E. Seneta. *Non-negative Matrices and Markov Chains*. Springer-Verlag, New York, 1981.
30. J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 15:733–736, 1963.
31. D. J. Hartfiel. *Markov set-chains*. Springer, Berlin;New York, 1998.
32. H. Tanner, A. Jadbabaie, and G. Pappas. Distributed coordination strategies for groups of mobile autonomous agents. Technical report, ESE Department, University of Pennsylvania, December 2002.
33. L. Moreau. Stability of multi-agent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, pages 169–182, February 2005.
34. W. Ren and R. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50:655–661, 2005.
35. D. Angeli and P. A. Bliman. Stability of leaderless multi-agent systems. 2004. technical report.
36. Z. Lin, B. Francis, and M. Brouche. Local control strategies for groups of mobile autonomous agents. *IEEE Trans. Auto. Control*, pages 622–629, april 2004.
37. V. D. Blondel, J. M. Hendrichx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. 2005. Submitted to IEEE CDC 2005.
38. R. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.
39. J. L. Doob. *Stochastic Processes*, chapter 5: Markov Processes, Discrete Parameter. John Wiley & Sons, Inc., New York, 1953.
40. J. N. Tsisiklis. *Problems in decentralized decision making and computation*. Ph.D dissertation, Department of Electrical Engineering and Computer Science, M.I.T., 1984.
41. C. F. Martin and W. P. Dayawansa. On the existence of a Lyapunov function for a family of switching systems. In *Proc. of the 35th Conf. on Decision and Contr.*, December 1996.
42. M. Cao, D. A. Spielman, and A. S. Morse. A lower bound on convergence of a distributed network consensus algorithm. In *Proc. 2005 IEEE CDC*, 2005.
43. B. Mohar. The Laplacian spectrum of graphs. *in Graph theory, combinatorics and applications (Ed. Y. Alavi G. Chartrand, O. R. Ollerman, and A. J. Schwenk)*, 2:871–898, 1991.
44. M. Cao, A. S. Morse, and B. D. O. Anderson. Coordination of an asynchronous multi-agent system via averaging. In *Proc. 2005 IFAC Congress*, 2005.
45. Miroslav Krstić, Ioannis Kanellakopoulos, and Petar Kokotović. *Nonlinear and Adaptive Control Design*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, New York, 1995.
46. Z. Lin, M. Brouche, and B. Francis. Local control strategies for groups of mobile autonomous agents. Ece control group report, University of Toronto, 2003.
47. A. Isidori. *Nonlinear Control Systems*. Springer-Verlag, 1989.