Topics in Dynamical Systems

Hook, JL

2012

MIMS EPrint: **2013.39**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

TOPICS IN DYNAMICAL SYSTEMS

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2012

James Hook

School of Mathematics

Contents

Word count xxxxx

The University of Manchester

James Hook
Doctor of Philosophy
Topics in Dynamical Systems
April 2, 2012

In this thesis I explore three new topics in Dynamical Systems. In Chapters 2 and 3 I investigate the dynamics of a family of asynchronous linear systems. These systems are of interest as models for asynchronous processes in economics and computer science and as novel ways to solve linear equations. I find a tight sandwich of bounds relating the Lyapunov exponents of these asynchronous systems to the eigenvalue of their synchronous counterparts. Using ideas from the theory of IFSs I show how the random behavior of these systems can be quickly sampled and go some way to characterizing the associated probability measures.

In Chapter 4 I consider another family of random linear dynamical system but this time over the Max-plus semi-ring. These models provide a linear way to model essentially non-linear queueing systems. I show how the topology of the queue network impacts on the dynamics, in particular I relate an eigenvalue of the adjacency matrix to the throughput of the queue.

In Chapter 5 I consider non-smooth systems which can be used to model a wide variety of physical systems in engineering as well as systems in control and computer science. I introduce the Moving Average Transformation which allows us to systematically 'smooth' these systems enabling us to apply standard techniques that rely on some smoothness, for example computing Lyapunov exponents from time series data.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made **only** in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of the School of Mathematics.

Acknowledgements

I have been very lucky and am extremely grateful for the mentoring I have received at the University of Manchester during my time as a PhD student. My supervisor David Broomhead's help and advice, his innovative work, friendship and enthusiasm have shaped the way I try to think about Mathematics. Mark Muldoon and Paul Glendining have also been very generous in their support, particularly in my final year.

I have always felt that these three doors are open to me whenever I am stuck with a problem or have something interesting I want to share and I am very grateful for this. Being able to work with Dave, Mark and Paul has not only been extremely beneficial through their invaluable suggestions but really good fun too!

There are many others in the CICADA group and the school of maths at large who have been good friends, taken an interest in my work or inspired me though their own. I would particularly like to thank my officemate Chris Welshman for many useful discussions.

I would also like to thank my parents Peter and Nicola Hook for their support throughout my education.

Chapter 1

Introduction

Preface

As a PhD student in the CICADA (Center for interdisciplinary computational and dynamical analysis) group I have been exposed to a wide variety of exciting new topics in applied mathematics. The group's focus on novel systems arising in Computer science, Control, Biology and Engineering has given me the opportunity to study a wide variety of so called *Hybrid systems*. Although there is no universally accepted definition of a Hybrid system any dynamical system which contains a mixture of discrete and continuous dynamics roughly fits the bill. The non-smooth systems I investigate in chapter 5 are certainly of this form and whilst the asynchronous processes considered in the earlier chapters are not hybrid as such they do share a common unusualness, in the sense that whilst having some tractable structure they do not easily fit into the standard theory of continuous or discrete dynamical systems.

Developing theory for these new systems which are of great interest in science and engineering is an important new challenge in mathematics and I hope to show that despite their apparent complexity many of these new systems can still be successfully investigated in an analytic way.

My Thesis consists of three main bodies of work; Asynchronous linear systems, Max-plus linear systems and Non-smooth systems which I present in chapters 2,3 and 4 and 5 respectively.

Chapters 2, 4 and 5 form the basis of research papers which I hope to submit shortly.

The remainder of this introductory chapter is divided along the lines of my three topics. Each chapter includes a fairly comprehensive introduction so the remaining sections of this first chapter simply introduce some of the basic concepts or highlights any pertinent standard theory not discussed in the following chapters.

1.1 Asynchronous linear systems

In chapters 2 and 3 our basic set up is as follows. We have a finite set of matrices $\{A(i) \in \mathbb{R}^{N\times N} : i = 1, 2, .., M\}$ which we randomly apply to some state $x \in \mathbb{R}^N$ by at each stage drawing $i_n \in \{1, 2, .., M\}$ with some Bernoulli or Markov rule and multiplying so that

$$x(n+1) = A(i_n)x(n) \tag{1.1}$$

The idea here is that each of the matrices corresponds to a different action in some system of agents with no global control over what happens. For instance we could have a group of agents trying to synchronize some value associated with them by randomly changing their value to be closer to their neighbors, without some hierarchy there is no way to control the order in which the agents update so we have such an asynchronous system where the matrix $A(i)$ corresponds to the ith agent updating their state.

For a class of systems with some additional structure I address the questions; i) when does this system converge? ii) if it exists is the equilibrium a random variable or somehow predetermined? and iii) what is the expected rate of convergence?

There are two important pieces of standard theory here. The Multiplicative ergodic theorem and the theory of Contraction on Average Iterated Function Systems (IFSs). Some key IFS results are quoted and referenced in chapter 2 sections 2.2 and 2.3. Several versions of the Multiplicative ergodic theorem have existed before Valery Oseledec's work but his is the strongest and most recent.

Definition $T : X \mapsto X$ is *ergodic* with ergodic measure μ if for all continuous $f : X \mapsto \mathbb{R}$ and μ almost all x

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f[T^k(x)] = \int_X f(x')d\mu \tag{1.2}$$

Definition A *cocyle* $C : X \times \mathbb{N} \mapsto \mathbb{R}^{n\times n}$ on a dynamical system $T : X \mapsto X$ is defined by, $C(x, 0) = I$ and

$$C(x, n + 1) = F[T^n(x)]C(x, n) \tag{1.3}$$

Where $F : X \mapsto \mathbb{R}^{n\times n}$ is continuous.

Theorem If $T : X \mapsto X$ is an ergodic map with ergodic measure μ and $C : X \times \mathbb{N} \mapsto \mathbb{R}^{n\times n}$ a cocycle on X then

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log \frac{||C(x, n)u||}{||u||} \tag{1.4}$$

exists for μ almost all x and depends only on u taking at most n different values, the Lyapunov exponents.

This result implies that if we choose matrices from a Markov chain and multiply them together the size of the real parts of the product's eigenvalues should scale with these Lyapunov exponents and therefore, with probability one, not depend on the sequence of matrices we choose.

1.2 Max-plus linear systems

In chapter 4 I investigate the dynamics of stochastically timed petri-nets (a fairly general queue type model) using Max-plus algebra. I fully detail the application, whose connection to the Max-plus system is quite involved, in the chapter so this section will serve simply to introduce the reader to the Max-plus algebra and Max-plus linear systems.

Formally the Max-plus algebra is the semi-ring $\mathbb{R}_{\max} = [\mathbb{R} \cup -\infty, \oplus, \otimes]$ where $a \oplus b = \max\{a, b\}$ and $a \otimes b = a + b$, we include minus infinity to act as the additive (taking the maximum) identity which we will denote ϵ throughout.

We can now define the moduloid $\mathbb{R}_{\max}^N$ by taking an N vector of elements from $\mathbb{R}_{\max}$ and further define linear maps on this moduloid in exact analogy with the linear maps on $\mathbb{R}^N$ with standard algebra, that is with matrices.

The map $T : \mathbb{R}_{\max}^N \mapsto \mathbb{R}_{\max}^N$ is linear if and only if there exists a max-plus matrix $M \in \mathbb{R}_{\max}^{N\times N}$ s.t. $T(x) = M \otimes x$ for all $x \in \mathbb{R}_{\max}^N$ where

$$[M \otimes x]_i = \bigoplus_{j=1}^{N} M_{i,j} \otimes x_j = \max_{j=1}^{N} M_{i,j} + x_j \tag{1.5}$$

So that linear maps on $\mathbb{R}_{\max}^N$ are exactly as the standard linear maps on $\mathbb{R}^N$ except we substitute max for plus and plus for times.

Deterministic, autonomous max-plus linear systems are therefore evolved by a dynamic of the form

$$x(n+1) = A \otimes x(n) = A^{\otimes n} \otimes x(0) \tag{1.6}$$

The matrix product $A^{\otimes n} = A \otimes A \otimes ... \otimes A$ has a very neat interpretation in terms of paths through weighted graphs. Suppose that G is a graph on N vertices with an edge from j to i of weight $A_{i,j}$ whenever $A_{i,j}$ is not equal to ϵ and no edge when this is the case. Now consider the components of $A \otimes A$ we have

$$[A \otimes A]_{i,j} = \max_{k=1}^{\infty} A_{i,k} A_{k,j} \tag{1.7}$$

which is the weight of the maximally weighted path of length 2 from j to i through the graph, likewise $A_{i,j}^{\otimes n}$ is the weight of the maximally weighted path of length n from j to i through the graph. Finally when we include the initial condition $x(0)$ we have the interpretation that $A^{\otimes n} \otimes x(0)_i$ is the weight of the maximally weighted path of length n that ends at i and begins with an initial weight $x(0)_j$ when the first vertex in the path is j.

For a stochastic or non-autonomous linear system we have

$$x(n+1) = A(n) \otimes x(n) = [\bigotimes_{k=1}^{n} A(k)] \otimes x(0) \tag{1.8}$$

for some sequence of Max-plus matrices $[A(k)]_{k=1}^{\infty}$. In this case we have more or less the same interpretation

$$[\bigotimes_{k=1}^{n} A(k)]_{i,j} = \max_{k_1,k_2,...,k_{n-1}} A(1)_{j,k_1} + A(2)_{k_1,k_2} + ... + A(n)_{k_{n-1},i} \tag{1.9}$$

is the weight of the maximally weighted path of length n from j to i that accumulates a weight from kth transition according to the components of the matrix $A(k)$ which we can think of as the maximally weighted path through a graph whose weights change from stage to stage.

In chapter 4 I show that an analogue of the multiplicative ergodic theorem holds for random Max-plus matrices and prove some new bounds on the Lyapunov exponents relating them to the topology of the graphs induced by the matrices.

1.3 Smoothing non-smooth systems

In chapter 5 I introduce an original technique for smoothing non-smooth systems using the Moving Average Transformation. In this introduction I shall briefly discuss via an example how my technique differs from the standard vector field regularization procedure.

Consider the non-smooth ODE

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 \\ \begin{cases} x & \text{for } y \geq 0 \\ 1 & \text{for } y < 0 \end{cases} \end{pmatrix} \tag{1.10}$$

By defining $\dot{x} = 1$ $\dot{y} = 0$ on $\{y = 0, x \leq 0\}$ we have existence and uniqueness of solution $\phi_t(x, y)$ for all positive time t for any initial condition (x, y).

This sliding system contains a sliding discontinuity on $\{y = 0, x \leq 0\}$ and a non-differentiable switch on $\{y = 0, x > 0\}$ as well as a grazing point at $(0, 0)$. See Fig 1,a.

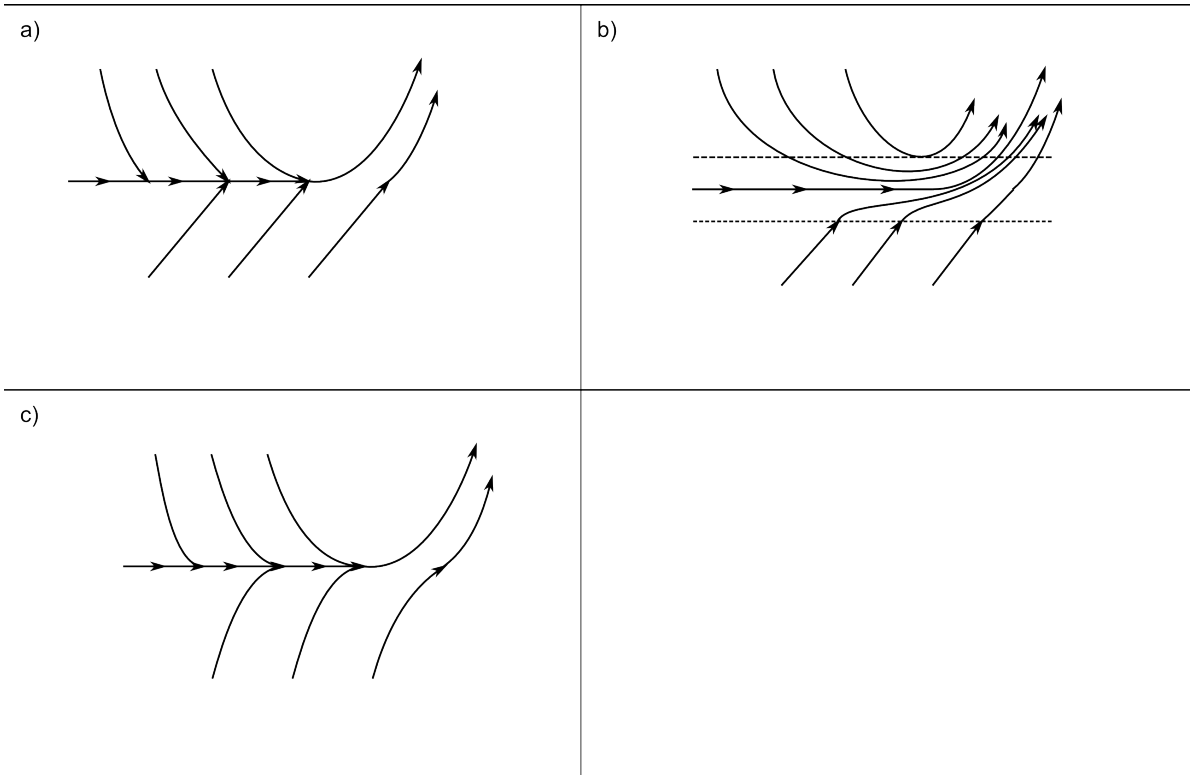


Figure 1.1: Sketch of solutions for a)Original non-differentiable system, b) Differentiable regularized system, c) Differentiable Moving Average Transformation transformed system.

The regularization procedure defines a new vector field

$$H(x,y) = h_\epsilon(y)[F(x,y) - G(x,y)] + G(x,y) \tag{1.11}$$

where F and G are the vector fields as defined above and below the switching surface respectively and

$$h_\epsilon(y) = \begin{cases} 0 & \text{for } y < 0 \\ \frac{1}{2} + \frac{y}{2\epsilon} & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases} \tag{1.12}$$

The new system $(\dot{x}, \dot{y}) = H(x,y)$ is differentiable and the sliding surface becomes the slow manifold of the resulting fast slow system. See Fig 1,b.

My technique employs the moving average transformation

$$(p,q) = \Phi(x,y) = \int_{-1}^{0} \phi_\tau(x,y) d\tau \tag{1.13}$$

which is multi-valued on $\{y = 0, x \leq 0\}$ as there is no unique backwards behavior from the sliding surface. This turns out not to matter, we simply take all the possible branches of Φ and define the transformed system

$$(\dot{p}, \dot{q}) = \begin{cases} \Phi'(p,q)F[\Phi^{-1}(p,q)] & \text{for } \Phi^{-1}(p,q)_2 \geq 0 \\ \Phi'(p,q)G[\Phi^{-1}(p,q)] & \text{for } \Phi^{-1}(p,q)_2 < 0 \end{cases} \tag{1.14}$$

which is differentiable. The sliding surface still 'absorbs' other orbits but in a differentiable manner through a square root term in the flow. See Fig 1,c.

Whilst the regularization approach is easier to understand and implement it changes the orbit structure of the system (the dynamics!) and results in a stiff system which can not be reliably investigated numerically.

My approach does not affect the orbit structure, indeed the original and transformed systems are topologically equivalent. Nor does it result in a stiff system as the action of the discontinuity is 'spread out' over the length of the averaging period which is not intended to be on a short scale.

In chapter 5 I rigourously investigate the action of the transformation on many of the typical non-smooth discontinuity normal forms and transform examples both analytically and numerically. As an application of the technique I calculate the Lyupanov spectrum of a chaotic non-smooth system from its time-series.

Chapter 2

Consensus in an asynchronous network: an IFS perspective

Abstract

We examine a linear model of a group of agents reaching consensus in an asynchronous networked environment. We show that the system can be described by an Iterated Function System IFS that contracts on average. This observation allows us to draw on a rich theory and in particular we make some quantitative links between the dynamics of the model and the topology of the network that it lives on.

Introduction

We are interested in asynchronous systems that live on networks. By this we mean that for some graph $G = \langle V, E \rangle$ the state of the whole system is the direct product of states at each vertex $x = \bigotimes_{v \in V} x_v$ and that associated with each vertex v is an 'update' function f_v whose application only changes the state of vertex v and only depends on the states of those vertices adjacent to v in G.

The state of our system is then evolved by applying a sequence of these functions to it. Asynchrony is reflected by the fact that the vertices do not update in unison. Instead they update in some random order and if we then choose our sequence of

updates $(\sigma_1, \sigma_2, ...)$ from some Markov chain then we arrive at a sequence of functions $(f_{\sigma_1}, f_{\sigma_2}, ...)$ which we can apply to the state to evolve the system. This is therefore a formalism with scope to define a wide variety of asynchronous agent based models [1]. For instance such a system could describe the dynamics of some series of interactions in a social network [2] or the evolution of a consensus or synchronization algorithm distributed over a large network of communicating robots [3]. In examining these systems it is natural to ask how the topology of the network impacts on the dynamics?

For the class of linear systems that we examine two useful theories are developed. One describes the rate of convergence to consensus. The other involves recasting a systems as an IFS then examining the Hausdorff dimension of the induced probability measure which it turns out is related to the topology of the network. Whilst this sounds very abstract it has a neat interpretation to do with the relative importance of the exact order in which vertices updates for a particular network.

This chapter is laid out as follows. In section 1 we introduce our linear model and prove that it is stable and that its rate of convergence is almost surely a constant which we bound with an expression involving the eigenvalue of an associated synchronous process. In section 2 we show how to recast the model as an IFS and for an explicit example use this formulation to quickly approximate statistics of the systems asymptotic behavior. In section 3 we use the commutivity relations defined by the network to bound the Hausdorff dimension of the measure associated with the IFS and treat some further examples.

2.1 The Model

Let $G = \langle V, E \rangle$ be an undirected connected finite graph. For each $v \in V$ we associate a state space $X_v = \mathbb{R}$ so that the systems state space is given by the direct product $X = \bigotimes_{v \in V} X_v = \mathbb{R}^{|V|=N}$. Next we assume that each vertex 'updates' itself independently of the others with exponential parameter λ_v inter-update-times. When a vertex v updates it does so by taking its new state to be a skewed mean of its old

state and the states of its neighbors so that

$$x_v \mapsto \pi_{v,v} x_v + \sum_{u:(v,u)\in E} \pi_{v,u} x_u \tag{2.1}$$

where $\pi_{v,v} + \sum_{u:(v,u)\in E} \pi_{v,u} = 1$ is a measure supported (i.e. non zero) on all indices in the sum.

So vertices independently choose to update at random times and when they do they average themselves to be more like the other vertices around them but not necessarily in a uniform way. That this system converges to a consensus when all vertices have the same state is fairly obvious. What is more interesting (hopefully!) is that although the value of this eventual equilibrium is a random variable the rate of convergence to it is almost surely a constant. The measure induced by this system (the probability distribution of the equilibria) can be very complicated, we show how to quickly approximate its statistics using backwards sampling and show how its fractal properties can be characterized by the topology of the graph $\mathcal{G}$.

In all subsequent examples we will use a less general formulation with $\pi_{i,j} = \frac{\epsilon}{d_i}$ where d_i is the degree of i in G and $\pi_{i,i} = 1 - \epsilon$ for some real parameter ϵ. We can therefore talk about the system on graph G without specifying the measures π.

2.1.1 Discrete Time Random Matrix Model

Index the vertices of G by $\{1, 2, .., N\}$. Updating vertex i is equivalent to multiplying the system state vector $x = (x_1, x_2, ..., x_N)^\top$ by the matrix $A(i)$ where

$$A(i)_{j,k} = \begin{cases} \pi_{i,i} & \text{for } j = k = i \\ \pi_{i,k} & \text{for } j = i,\ (j,k) \in E \\ 1 & \text{for } j = k \neq i \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

So that $A(i)$ is the identity except for the ith row which takes its entries from the measure π. Now the vertices update independently and randomly with exponential parameter β_i inter-update-times and this induces a measure on the possible sequences of updates. Let $S = \{1, 2, ..., N\}^{\mathbb{N}}$ be the set of infinite sequences of vertices so that

an element $\sigma = (\sigma_1, \sigma_2, ...)$ corresponds to σ_1 updating first then σ_2 and so on. At any point in time the probability that the ith vertex is the next to update is given by

$$\mu_i = \frac{\frac{1}{\beta_i}}{\sum_{j=1}^{N} \frac{1}{\beta_j}} \tag{2.3}$$

So the measure induced on S is Bernoulli $(\mu_1, \mu_2, ..., \mu_N)$. We can now define the discrete time system

$$x(n) = A(\sigma_n)A(\sigma_{n-1})...A(\sigma_1)x(0) = \mathcal{A}(\sigma)|_n[x(0)] \tag{2.4}$$

Where $\mathcal{A}$ takes a sequence in $\{1, 2, ..., N\}$ to the corresponding product of matrices which we then apply to the state. This is equivalent to the original continuous time system since the nth stage of the discrete time system is exactly the state of the continuous time system at time $T(n)$ the time of the nth update. We will make use of *the law of large numbers* throughout this chapter so we state it here for completeness

Theorem Suppose that $(x_n)_{n=1}^{\infty}$ is a sequence of i.i.d. random variables with finite expectation, $\mathbb{E}x_i = \mu < \infty$ then

$$\mathbb{P}[\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} x_k - \mu = 0] = 1 \tag{2.5}$$

Which implies

$$\lim_{n\to\infty} \frac{T(n)}{n} = \frac{1}{\sum_{j=1}^{N} \beta_j} \tag{2.6}$$

which we will use to reconcile our convergence results for the discrete system with the continuous time system.

2.1.2 Convergence to Equilibrium

All of the $A(i)$ are row-stochastic so they act as the identity on the diagonal

$$D = \{x \in \mathbb{R}^d : x_1 = x_2 = ... = x_N\} \tag{2.7}$$

Theorem With probability-1 $\lim_{n\to\infty} x(n)$ exists and is in D. We will write $\lim_{n\to\infty} x(n) = d[x(0), \sigma]$, since the eventual equilibrium is a well defined function of the initial condition and update sequence.

Proof Define $t(x) \in \{1, 2, ..., N\}$ s.t. $x_{t(x)} \geq x_i \ \forall i$ and likewise $b(x) \in \{1, 2, ..., N\}$ s.t. $x_{b(x)} \leq x_i \ \forall i$ choosing the greatest possible $t(x)$ and least $b(x)$ if they are non-unique. Now $\delta_D(x) = x_{t(x)} - x_{b(x)} = 0$ if and only if $x \in D$. We shall show that with probability 1 $\lim_{n\to\infty} \delta_D(x(n)) = 0$.

Claim 1 $\delta_D[x(n)]$ is non-increasing

Proof Suppose that $x_i = x_{t(x)}$ and that we apply the map $A(i)$ the new value of x_i will be $\sum_{j=1}^{N} A(i)_{i,j} x_j \leq x_i$ so $x_{t(x)}$ can not increase and likewise $x_{b(x)}$ can not decrease so $\delta_D(x)$ is non-increasing.

Now suppose that $x_{i_1} = x_{i_2} = ... = x_{i_m} = x_{t(x)}$ we shall show that there exists a finite sequence of updates that will decrease $\delta_D(x)$ and that the probability associated with the sequence and the fraction that it decreases δ_D can both be bounded.

Construction of ς For each i_j let $b(x), \varsigma^j(1), \varsigma^j(2), ..., (\varsigma^j(m) = i_j)$ be the shortest path from $b(x)$ to i_j in G - a sequence of vertices in G. Now define $\varsigma = \varsigma^m \varsigma^{m-1} ... \varsigma^1$ to be the sequence obtained by sticking all the ς^j together.

Claim 2 $\delta_D[\mathcal{A}(\varsigma)(x)] \leq \theta \delta_D(x)$ for some fixed $\theta < 1$

Proof We assume the 'worst case scenario' in which all vertices but $b(x)$ have the value $x_{t(x)}$. Now suppose we apply the sequence ς^j. Since $\varsigma^j(1)$ is connected to $b(x)$ we have

$$x(\varsigma^j(1)) \mapsto x_{t(x)}(1 - A_{b(x),\varsigma^j(1)}) + x_{b(x)} A_{b(x),\varsigma^j(1)} \tag{2.8}$$

and $\varsigma^j(2)$ is connected to $\varsigma^j(1)$ so

$$x(\varsigma^j(2)) \mapsto x_{t(x)}(1 - A_{b(x),\varsigma^j(1)} A_{\varsigma^j(1),\varsigma^j(2)}) + x_{b(x)} A_{b(x),\varsigma^j(1)} A_{\varsigma^j(1),\varsigma^j(2)} \tag{2.9}$$

and so on until finally

$$x(i_j) \mapsto x_{t(x)} \underbrace{(1 - A_{\varsigma^j(m-1),\varsigma^j(m)} ... A_{\varsigma^j(b(x)),\varsigma^j(1)})} + x_{b(x)} A_{\varsigma^j(m-1),\varsigma^j(m)} ... A_{b(x),\varsigma^j(1)} \tag{2.10}$$

$= \gamma_{b(x),i_j}$ and since we assumed the worst case scenario

$$\delta_D[\mathcal{A}(\sigma)(x)] \leq \sup_j \gamma_{b(x),i_j}\delta_D(x) \leq \sup_{i,j} \gamma_{i,j}\delta_D(x) \tag{2.11}$$

$= \gamma d_D(x)$

Claim 3 $\mu[\varsigma] \geq \rho$ for some fixed $\rho > 0$

Proof $\mu[\varsigma] = \prod_{j=1}^{m} \mu[\varsigma^j] \geq \inf_j \mu[\varsigma^j]^m$ and since we can have at most $N-1$ vertices with the state $t(x)$ we have $\mu[\varsigma] \geq \inf_j \mu[\varsigma^j]^{d-1}$. Now $\inf_j \mu[\varsigma^j]$ is the least probability associated with a path from $b(x)$ to one of our i_j which is bounded by the least probable shortest-path between any two vertices $\widetilde{\rho}$ so that

$$\mu[\varsigma] \geq \widetilde{\rho}^{d-1} = \rho \tag{2.12}$$

So at step n with probability greater than ρ

$$\delta_D[x(n+k)] \leq \theta\delta_D[x(n)] \tag{2.13}$$

where $k < (d-1)^2$ is the maximum length of ς and with probability less than $(1-\rho)$

$$\delta_D[x(n+k)] \leq \delta_D[x(n)] \tag{2.14}$$

therefore by the law of large numbers

$$\lim_{nk\to\infty} \log \delta_D[x(nk)] \to -\infty \tag{2.15}$$

so with probability 1 $\lim_{n\to\infty} \delta_D[x(n)] = 0$.

Claim 4 $\lim_{m\to\infty} |x(n) - x(n+m)| \leq C\delta_D[x(n)]$

Proof We have already shown that $x(t(x))$ is non-increasing and $x(b(x))$ is non-decreasing so for each vertex i we have $x_{b(x(n))} \leq x(n+m)_i \leq x_{t(x(n))}$ so that in the l_1 norm we have

$$\lim_{m\to\infty} |x(n) - x(n+m)| \leq N \times \delta_D[x(n)] \tag{2.16}$$

and we can find another constant for any equivalent norm.

Therefore as the system approaches D it slows down in all directions and really does converge to an equilibrium point $d[\sigma, x(0)]$.

So our system converges to a point on the diagonal D which is a function of the sequence of updates so ought to be thought of as a random variable.

2.1.3 Rate of Convergence

The rate of convergence of $x(n) = \mathcal{A}(\sigma)|_n[x(0)]$ is determined by the second largest eigenvalue of the matrix $\mathcal{A}(\sigma)|_n$ which is not ideal so we modify the system as follows. Set $B(i) = PA(i)$ where P is an orthogonal projection onto the hyperplane

$$D^{\perp} = \{x \in \mathbb{R}^d : x_1 + x_2 + ... + x_N = 0\} \tag{2.17}$$

now for each n

$$x(n) = \mathcal{A}(\sigma)|_n x(0) = B(\sigma_n)...B(\sigma_1)x(0) + d[n, \sigma, x(0)] = \mathcal{B}(\sigma)|_n x(0) + d[n, \sigma, x(0)] \tag{2.18}$$

for some function $d : \mathbb{N} \times S \times \mathbb{R}^n \to D$ which records the information lost by the repeated application of the projection. The rate of convergence of the system is now determined by the greatest eigenvalue of $\mathcal{B}(\sigma)|_n$ which we denote $\mathcal{X}[\mathcal{B}(\sigma)|_n]$.

Theorem With probability-1 $\lim_{n\to\infty} \frac{1}{n} \log \mathcal{X}[\mathcal{B}(\sigma)|_n] = \lambda$ exists and is a constant independent of the sequence of updates.

Proof Define the modified geometric means λ_n by

$$\lambda_n = \sum_{\sigma|_n} \frac{\mu[\sigma|_n]}{n} \log \mathcal{X}[\mathcal{B}(\sigma)|_n] \tag{2.19}$$

where the sum is taken over all the distinct sequences of length n.

Claim 1 λ_n is non-increasing and converges

Proof Consider

$$\lambda_{n+m} = \sum_{\sigma|_n} \sum_{\varsigma|_m} \frac{\mu[\sigma|_n]\mu[\varsigma|_m]}{n+m} \log \mathcal{X}[\mathcal{B}(\sigma)|_n \mathcal{B}(\varsigma)_m] \tag{2.20}$$

using the fact that $\mathcal{X}(A)\mathcal{X}(B) \geq \mathcal{X}(AB)$ and summing over each set of sequences we obtain

$$(n+m)\lambda_{n+m} \leq n\lambda_n + m\lambda_m \tag{2.21}$$

Lemma Fekete's Subadditive Lemma states that for a subadditive sequence, that is a sequence $[a(n) \in \mathbb{R}]_{n=1}^{\infty}$ with $a(n+m) \leq a(n) + a(m)$ the limit $\lim_{n\to\infty} \frac{a(n)}{n} = \sup_{n\to\infty} \frac{a(n)}{n}$ exists. so by Fekete's Subadditive Lemma $(\lambda_n)_{n=1}^{\infty}$ is a decreasing sequence that converges to some constant λ.

Now suppose that we have some long sequence $\sigma|_{nm+k}$ with $k < m, n$ which we decompose as $(\sigma^0|_k\sigma^1|_n\sigma^2|_n...\sigma^m|_n)$ then

$$\frac{1}{nm+k}\log\mathcal{X}[\mathcal{B}(\sigma)|_{nm+k}] \leq \frac{1}{nm+k}(\log\mathcal{X}[\mathcal{B}(\sigma^0)|_k] + \sum_{i=1}^{m}\log\mathcal{X}[\mathcal{B}(\sigma^i)|_n]) \tag{2.22}$$

and by the law of large numbers we have $\lim_{m\to\infty}\frac{1}{nm+k}\log\mathcal{X}[\mathcal{B}(\sigma)|_{nm+k}] \leq \lambda_n$ and from (2.20) $\lim_{m,n\to\infty}\lambda_n = \lambda_{nm+k}$.

Claim 2 With probability 1 $\lim_{m,n\to\infty}\frac{1}{nm+k}\log\mathcal{X}[\mathcal{B}(\sigma)|_{nm+k}] = \lambda_{nm+k} = \lambda$

Proof Suppose that $\mu[\{\sigma|_{nm+k} : \frac{1}{nm+k}\log\mathcal{X}[\mathcal{B}(\sigma)|_{nm+k}] < \lambda_{nm+k} - \epsilon\}] = p > 0$ then since λ_{nm+k} is the mean of these quantities

$$\lambda_{nm+k} \leq (1-p)\lambda_{nm+k} + p(\lambda_{nm+k} - \epsilon) \tag{2.23}$$

which is a contradiction.

So with probability-1 $\lim_{n\to\infty}\frac{1}{n}\log\mathcal{X}[\mathcal{B}(\sigma)|_n] = \lambda$ where $\lambda = \lim_{n\to\infty}\lambda_n$.

Remark This proof has provided us with a series of arbitrarily tight upper bounds to λ the λ_n. However these are ridiculously expensive to compute since we would have to sum over d^n sequences to obtain λ_n. We have already used this idea, in section 1 we showed that $\lambda \leq \lambda_k < 1$ which was enough to prove that the system converges to an equilibrium.

To obtain a more useful upper bound on λ we must compare our asynchronous process to a corresponding synchronous process. Define the synchronous matrix A by

$$A = [\sum_{i=1}^{N} A(i)] - (N-1)I \tag{2.24}$$

So that multiplying a state by A corresponds to all the vertices simultaneously updating their states. A is non-negative irreducible and aperiodic.

Upper bound $\lambda \leq \rho + \frac{1}{2}(1-\frac{1}{N})\widehat{\rho}$ Where

$$\rho = \sup_{y:\sum_i y_i=0} \frac{1}{\langle y,y\rangle} y^T \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_N \end{pmatrix} (A-I)y \tag{2.25}$$

and

$$\widehat{\rho} = \sup_{y:\sum_i y_i=0} \frac{1}{\langle y,y\rangle} [(A-I)y]^T \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_N \end{pmatrix} (A-I)y \tag{2.26}$$

In the case that the update rates are uniform, i.e. $\mu_1 = \mu_2 = ... = \mu_N = \frac{1}{N}$ we have $\sqrt{\widehat{\rho}} = \rho = \varrho - 1$ where ϱ is the second largest eigenvalue of A.

Proof Since λ is almost surely a constant we can express it as an expectation

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{\sigma|_n} \mu[\sigma|_n] \log H[\mathcal{A}(\sigma|_n)x] \tag{2.27}$$

Which is independent of x and where H is the semi-norm measuring the l_2 distance to the diagonal D.

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{\sigma|_{m-1}} \mu[\sigma|_{m-1}] \underbrace{(\sum_{i=1}^{N} \mu_i \log H[A(i)\mathcal{A}(\sigma|_{m-1})x] - \log H[\mathcal{A}(\sigma|_{m-1})x])}$$

and the underbraced is less than or equal to

$$\sup_{y:H(y)=1,\sum_i y_i=0} \sum_{i=1}^{N} \mu_i \log H[A(i)y] \tag{2.28}$$

Now

$$H[A(i)y]^2 = [y_i + \langle L_i, y\rangle - \frac{1}{N}\langle L_i, y\rangle]^2 + \sum_{j\neq i} [y_j - \frac{1}{N}\langle L_i, y\rangle]^2 \tag{2.29}$$

Where L_i is the ith row of $(A - I)$.

$$H[A(i)y]^2 = \sum_{j=1}^{N} y_j^2 - \frac{2\langle L_i, y\rangle}{N} \sum_{j=1}^{N} y_j + 2y_i\langle L_i, y\rangle + (1 - \frac{1}{N})\langle L_i, y\rangle^2 \quad (2.30)$$

$$= 1 + 2y_i\langle L_i, y\rangle + (1 - \frac{1}{N})\langle L_i, y\rangle^2$$

Where we have used the face that $\sum_{i=1}^{N} y_i = 0$ and $\sum_{i=1}^{N} y_i^2 = 1$. Using $\log(1 + x) \leq x$ the underbraced term is less than or equal to

$$\sup_{y:H(y)=1,\sum_{i=1}^{N} y_i=0} \frac{1}{2} \sum_i \mu_i[2y_i\langle L_i, y\rangle + (1 - \frac{1}{N})\langle L_i, y\rangle^2] \quad (2.31)$$

$$= \sup_{y:H(y)=1,\sum_i y_i=0} [y + \frac{1}{2}(1 - \frac{1}{N})(A - I)y]^T \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_N \end{pmatrix} (A - I)y$$

$$= \rho + \frac{1}{2}(1 - \frac{1}{N})\widehat{\rho}$$

which on substitution back into (35) yields

$$\lambda \leq \rho + \frac{1}{2}(1 - \frac{1}{N})\widehat{\rho} \quad (2.32)$$

which in the uniform update rate case gives

$$N\lambda \leq \varrho - 1 + \frac{1}{2}(1 - \frac{1}{N})(\varrho - 1)^2 \quad (2.33)$$

2.1.4 Example

Let $\mathcal{G}$ be the complete graph on three vertices. If the update function of each vertex v with neighbors u, w is the skewed mean

$$f_v(x_v; x_u, x_w) = (1 - \epsilon)x_v + \frac{\epsilon}{2}(x_u + x_w) \quad (2.34)$$

then the random matrix product system consists of the three matrices

$$A(1) = \begin{pmatrix} 1 - \epsilon & \frac{\epsilon}{2} & \frac{\epsilon}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad A(2) = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\epsilon}{2} & 1 - \epsilon & \frac{\epsilon}{2} \\ 0 & 0 & 1 \end{pmatrix} \quad A(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\epsilon}{2} & \frac{\epsilon}{2} & 1 - \epsilon \end{pmatrix} \quad (2.35)$$

which we apply according to a bernoulli measure $\mu = (\mu_1, \mu_2, \mu_3)$. For μ uniform we compare an accurate numerical estimate of λ to our upper bound over a range of ϵ. We carry out the same simulation for the system on the complete graph of twenty vertices.

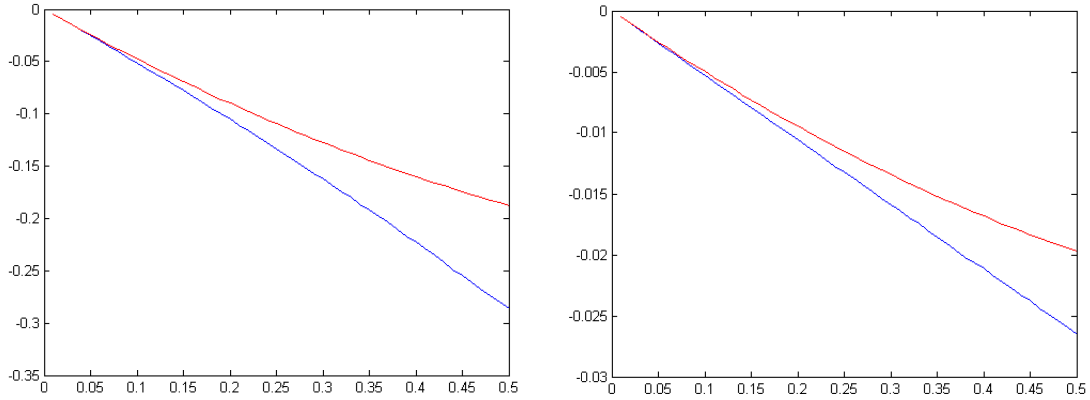


Figure 2.1: λ (blue) and upper bound (red) vs ϵ for $N = 3$ (left) and $N = 20$ (right).

2.2 IFS Perspective

We can recast out system as an IFS. From section 1 we know that

$$\lim_{n\to\infty} A(\sigma(n))...A(\sigma(1))x(0) = \alpha(\sigma, x(0))(1, ..., 1)^\top \tag{2.36}$$

Where $\alpha(\sigma, x(0))$ is the eventual consensus value which is some unknown function of the initial condition and the update sequence. Multiplying both sides on the right by $\frac{1}{N}(1, ..., 1)$ we obtain

$$\lim_{n\to\infty} \langle A(\sigma(n))...A(\sigma(1))x(0), \frac{1}{N}(1, ..., 1)^\top \rangle = \alpha(\sigma, x(0)) \tag{2.37}$$

We can now take the $A(i)$ to the other side of the inner product so that

$$\lim_{n\to\infty} \langle x(0), \underbrace{A^\top(\sigma(1))...A^\top(\sigma(n))\frac{1}{N}(1, ..., 1)^\top}\rangle = \alpha(\sigma, x(0)) \tag{2.38}$$

$= \phi(\sigma|_n)$. The value of the consensus can therefore be expressed as the scalar product of the initial condition and the limit of a new system which only depends on the sequence σ.

Definition An *Iterated Function System* is a triple $[X, F, \mu]$ where X is some state space, F is a finite set of functions $f_i : X \to X$ and μ is a bernoulli measure on the set of infinite sequences of functions in F. We have two process associated with the system

$$\varphi|_n(x, \sigma) = f_{\sigma(n)} \circ ... \circ f_{\sigma(1)}(x) \tag{2.39}$$

where σ is a μ distributed random variable is the *forwards process* and

$$\phi|_n(x, \sigma) = f_{\sigma(1)} \circ ... \circ f_{\sigma(n)}(x) \tag{2.40}$$

is the *backwards process.*

Definition We say that an IFS is a *Contraction on Average* if

$$\lambda' = \int_\sigma \lim_{n\to\infty} \frac{1}{n} \sup_{x,y} \log \frac{|\varphi_n(x, \sigma) - \varphi|_n(y, \sigma)|}{|x - y|} d\sigma \tag{2.41}$$

is less than zero.

Theorem If $[X, F, \mu]$ is a Contraction on Average IFS then

- $\lim_{n\to\infty} \phi|_n(x, \sigma)$ exists and is independent of x
- The measure

$$\nu_F(A \in X) = \lim_{n\to\infty} \frac{1}{n} \sum \mathbb{I}(\varphi|_n(x, \sigma) \in A) \tag{2.42}$$

  Is well defined and for a μ measure one set identical to the measure

$$\nu_B(A \in X) = \mu\{\sigma : \lim_{n\to\infty} \phi|_n(x, \sigma) \in A\} \tag{2.43}$$

Proof See Diaconis and Freedman [4]

Theorem $\phi|_n(\sigma)$ as defined in (2.39) gives rise to a contraction on average IFS on the unit simplex.

Proof The unit simplex is defined by

$$\Delta_d = \{x \in \mathbb{R}^d : x_1 + x_2 + ... + x_d = 1 : x_j \geq 0 \forall j\} \tag{2.44}$$

Recall that

$$\phi(\sigma)|_n = A^T(\sigma(1))...A^T(\sigma(n))\frac{1}{d}(1,...,1)^t \tag{2.45}$$

now the $A(i)$ are non-negative and row stochastic so $A(i)^T : \Delta_n \to \Delta_n$. All that remains is to show that we have a contraction on average which we shall do using the following metric on $\Delta|_n$

$$d(x,y) = \sup_{z\in\Delta|_n} |\langle x-y, z\rangle| \tag{2.46}$$

Now the integrand of (2.40) becomes

$$\lim_{n\to\infty}\frac{1}{n}\sup_{x,y\in\Delta|_n}\log\frac{\sup_{z\in\Delta|_n}|\langle\varphi_n(x,\sigma)-\varphi_n(y,\sigma),z\rangle|}{\sup_{z'\in\Delta|_n}|\langle x-y,z'\rangle|} \tag{2.47}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sup_{x,y\in\Delta|_n}\log\sup_{z\in\Delta|_n}|\langle x-y,\mathcal{A}(\sigma|_n)z\rangle|$$

And with probability one $\mathcal{A}(\sigma|_n)z = d(z,\sigma) + \exp(\lambda n)d^{\perp}(n,z,\sigma)$ where $d \in D$ and $d^{\perp}$ is perpendicular to D and does not grow or decay exponentially in size. Therefore for almost all σ the integrand equals

$$\lambda + \lim_{n\to\infty}\frac{1}{n}\sup_{x,y\in\Delta|_n}\log\sup_{z\in\Delta|_n}|\langle x-y,d^{\perp}(n,z,\sigma)\rangle = \lambda \tag{2.48}$$

So that $\lambda' = \lambda < 0$ as claimed.

2.2.1 Backwards sampling

Now suppose that we want to know the probability that we converge to a particular set $K \subset D$ from some initial condition x. A naive approach would be to simulate the process many times and record the fraction of times we arrived in K this would clearly take a very long time to give us an accurate answer. Instead we can easily find the set $\widetilde{K} \subset \Delta_d$ s.t. $\lim_{n\to\infty}\phi|_n(x,\sigma) \in \widetilde{K}$ if and only if $\lim_{n\to\infty} x(n) \in K$ that is the set in the IFS system $\widetilde{K}$ corresponding to a set of equilibrium points in the original system K. Now we compute $\nu_F(\widetilde{K})$ from a simulation of the forwards process. The number we obtain will be the probability we are looking for only now we essentially obtain a statistically correct trial run of the system in every step of the calculation vastly improving the rate of convergence. We can use this trick to approximate any statistic of ν that we like.

Example

Using the sampling method outlined we can quickly approximate the distribution on equilibria for a given initial condition. Using the network in Fig 2 with the randomly generated initial condition

$$x = (0.29, 0.93, 0.55, 0.74, 0.65, 0.91, 0.53, 0.97) \tag{2.49}$$

we record the sequence $\langle x, \varphi|_n(\sigma)\rangle$ for 10000 iterations and use this data to produce the distributions approximation with mean 0.68 and standard deviation 0.028 shown in Fig 3.

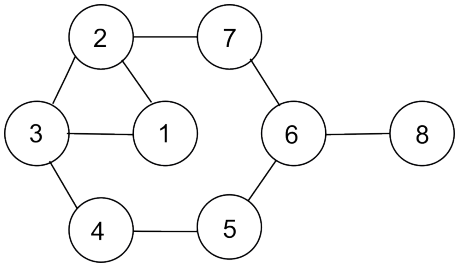


Figure 2.2: Simulated network.

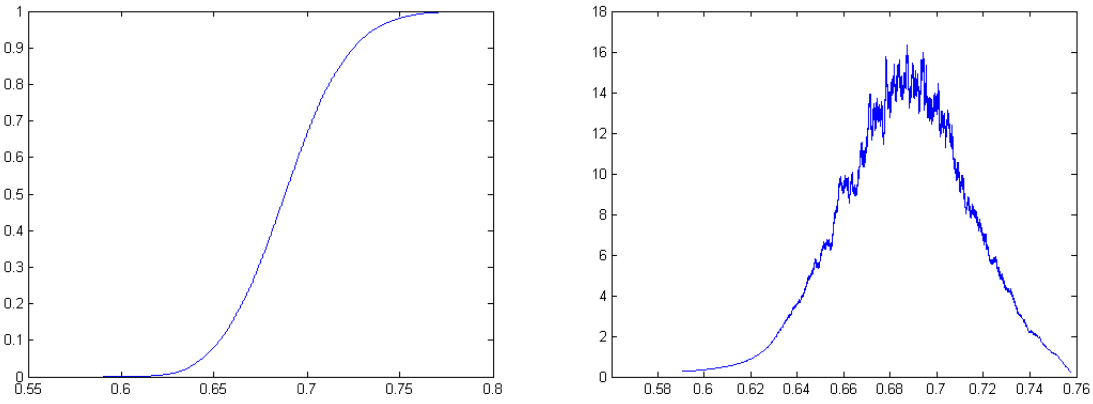


Figure 2.3: CDF and PDF for sampled data.

2.2.2 Barycentric form

The $A^T(i)$ that constitute our IFS act on Δ_N as affine maps (since the maps are linear on $\mathbb{R}^N$ and Δ_N is an affine subspace). Conveniently they admit the same standard matrix representation when we use the slightly redundant (in the sense that we have one more co-ordinate than dimension) barycentric co-ordinate system defined on Δ_N as follows.

The boundary of Δ_N consists of d copies of Δ_{d-1} which we can enumerate by

$$\delta(\Delta_N)_i = \{x \in \mathbb{R}^d : x_1 + x_2 + ... + x_N = 1 : x_j \geq 0 \forall j : x_i = 0\} \tag{2.50}$$

The barycentric co-ordinates of a point $x \in \Delta_N$ are then given by $x_i = d[x, \delta(\Delta_N)_i]$ where $[x, \delta(\Delta_N)_i]$ is the distance from x to the ith boundary of Δ_N.

2.2.3 Example

For the same system treated in 1.4 the IFS lives on Δ_3 which is an equilateral triangle embedded in $\mathbb{R}^3$.

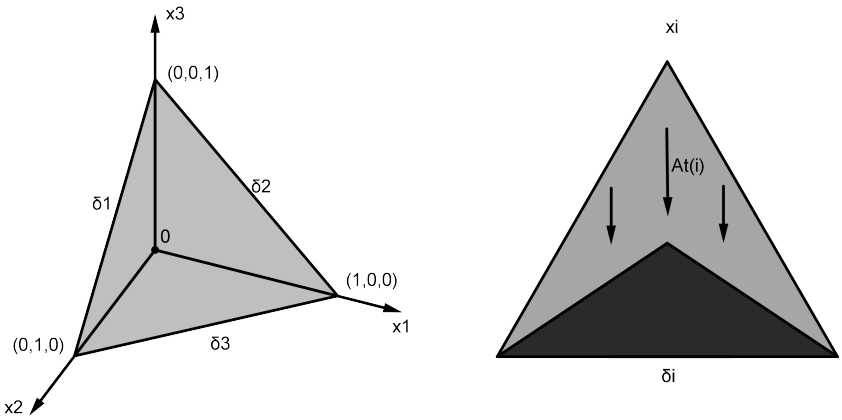


Figure 2.4: Orientation of Δ_3 in $\mathbb{R}^3$ and action of $A^\top(i)$.

The action of the map $A^T(i)$ is to contract the simplex towards the ith part of its boundary, equivalently we are pushed away from the ith co-ordinate axis. This means that when we eventually take the scalar product with our initial condition the

ith state is less important as a result of updating - which makes sense because when a vertex updates it is attempting to average itself into the group. The measure ν induced by this system can be very complicated.

- For $\epsilon \leq \frac{2}{3}$ ν appears to be supported continuously with a distribution whose mass becomes more and more centered as ϵ is made smaller.

- For $\epsilon = \frac{2}{3}$ it is easy to show that ν is uniform on Δ_3. For an initial condition $x(0) = (0, 1, a \in [0, 1])$ this induces a density on the set of possible consensus values given by

$$\omega(c) = \begin{cases} \frac{c}{2a} & \text{for } c \leq a \\ \frac{1-c}{2(1-a)} & \text{otherwise} \end{cases} \tag{2.51}$$

  which can easily be modified for an arbitrary initial condition by translation, dilation and some permutation.

- For $\epsilon > \frac{2}{3}$ ν is supported on a fractal set. That is there exist sets of ν measure 1 but zero Lebesgue measure (area). By recording an orbit of the foreword process we can obtain a scattering of points from one of these sets. For $\epsilon = 0.75$ we simulate such an orbit see Fig 5.

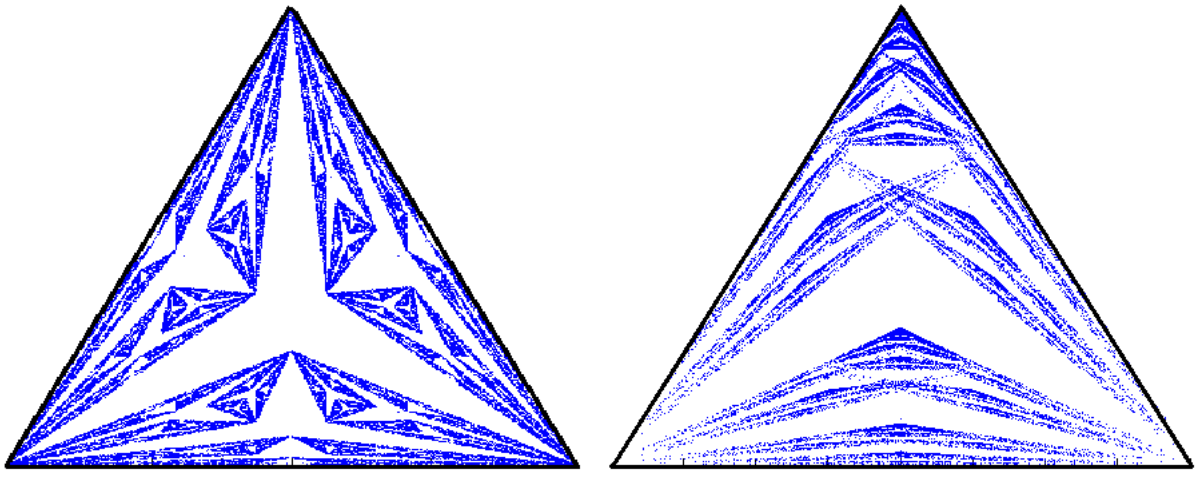


Figure 2.5: Sampled orbit from system on complete graph of three vertices (left) and a graph of three vertices in a line (right).

We can do the same thing for the network of three vertices on a line. The matrix product system consists of the following matrices

$$A(1) = \begin{pmatrix} 1-\epsilon & \epsilon & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad A(2) = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\epsilon}{2} & 1-\epsilon & \frac{\epsilon}{2} \\ 0 & 0 & 1 \end{pmatrix} \quad A(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \epsilon & 1-\epsilon \end{pmatrix} \tag{2.52}$$

Now vertices 1 and 3 are not connected so that the result of applying f_1 does not depend on the state of vertex 3 and visa versa. Thus the action of the map $A^\top(1)$ on Δ_3 is still to contract towards the 1st part of the boundary but in a way that conserves the distance from the 3rd part. Again we simulate the foreword process this time for $\epsilon = 0.6$.

In general if $\mathcal{G}$ is a graph on N vertices and U_i is the set of vertices in G not connected to vertex i then $A^\top(i)$ acts by contracting Δ_N towards $\delta[\Delta_N]_i$ in such a way as to preserve the distance from each $\delta[\Delta_N]_j : j \in U_i$.

2.3 Characterizing ν

As we have seen the induced measure ν can be very complicated. We can characterize this complexity with a fractal dimension which we first define for a set X. Suppose that $N(\epsilon)$ is the number of diameter ϵ sets required to cover X then

$$\dim_{BOX}(X) = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{-\log \epsilon} \tag{2.53}$$

this we call the box counting dimension. This notion of dimension coincides with the conventional meaning of dimension when we consider simple objects like manifolds. The dimension of a measure is then defined by

$$\dim_{BOX}(\nu) = \inf\{\dim_{BOX}(X) : \nu(X) = 1\} \tag{2.54}$$

Hausdorff dimension is another notion of fractal dimension with some nice additional properties [5] and although we will not use it directly we keep in mind the inequality

$$\dim_{BOX} \geq \dim_H \tag{2.55}$$

and since we will bound the box dimension above we automatically get the same result for the Hausdorff dimension.

Broomhead et al. show that the dimension of the measure induced by a contraction on average IFS can be bounded by the entropy of the symbol sequence (the random sequence of updates) divided by the Lypanov exponent of the system [6]. In this section we sketch a proof of their result for our special case and show how to calculate the entropy of the symbol sequence from the topology of the network.

2.3.1 Robust commutivity relations

We say that a set of functions $\{f_i : \mathbb{R}^d \mapsto \mathbb{R}^d\}$ satisfy a commutivity relation if an identity of the form

$$f_i \circ f_j(x) = f_j \circ f_i(x) \tag{2.56}$$

holds for all x and some i, j

Theorem Commutivity relations are not Typical

Proof We must first define what we mean by typical. Let $\mathcal{F}$ be the set of all pairs of continuous functions $f_1, f_2 : \mathbb{R}^d \mapsto \mathbb{R}^d$ and equip $\mathcal{F}$ with the metric $d[(f_1, f2), (g_1, g_2)] = \max\{\sup_{x\in\mathbb{R}^d}(f_1(x) - g_1(x)), \sup_{x\in\mathbb{R}^d}(f_2(x) - g_2(x))\}$. Now define C to be the set of all pairs that commute

Claim $\overline{C}$ is an open and dense subset of $\mathcal{F}$

Proof First we show that $\overline{C}$ is open. Consider the commutator map $\pi_c(f, g) = f \circ g - g \circ f$ which is clearly continuous and $\pi_c\overline{C} = \pi_c\mathcal{F}/\{0\}$ is open in $\pi_c\mathcal{F}$ so $\overline{C}$ must be open in $\mathcal{F}$. To show that $\overline{C}$ is dense consider a pair of points (f_1, f_2) which commute. Now chose any fixed x_0 to define the perturbed pair

$$g_1(x) = \begin{cases} f_1 \circ f_2(x_0) + \underline{\epsilon} & \text{for } x = f_2(x_0) \\ f_1(x) & \text{otherwise} \end{cases} \tag{2.57}$$

$$g_2 = f_2$$

Now $g_1 \circ g_2(x_0) = f_1 \circ f_2(x_0) + \underline{\epsilon}$ and $g_2 \circ g_1(x_0) = f_1 \circ f_2(x_0)$ unless we also have $f_2(x_0) = x_0$ and $f_2[f_1(x_0) + \underline{\epsilon}] = f_2 \circ f_1(x_0) + \underline{\epsilon}$ in which case we make the second perturbation

$$h_1 = f_1$$
$$h_2(x) = \begin{cases} g_2(x_0 + \underline{\epsilon}) + \underline{\epsilon} & \text{for } x = x_0 + \underline{\epsilon} \\ g_2(x) & \text{otherwise} \end{cases} \tag{2.58}$$

And obtain a non commuting pair within an $\|\underline{\epsilon}\|$ distance of our original commuting pair.

Remark The same result follows when we take $\mathcal{F}$ to be the set of pairs functions that are continuous, differentiable, twice differentiable, ..., smooth.

A possible interpretation of this result is that we should be very cautious of any mathematical analysis of a model for a physical system that relies on exploiting a commutivity relation. This is because the physical reality of the system that we are modeled can be thought of as a perturbation of our mathematical model and according to the theorem a generic perturbation will destroy any commutivity relations.

Definition An IFS $[X, F, \mu]$ is $G = \langle V, E \rangle$ *admissible* if there exists a bijection between the vertices of G and the functions constituting the IFS so that they can be written $F = \{f_v : v \in V\}$ and f_u and f_v commute whenever u and v are non adjacent in G.

Theorem The commutivity relations satisfied by functions $\{f_v : v \in V\}$ associated with a $G = \langle V, E \rangle$ admissible IFS are robust in the sense that they persist for any perturbation to the f_i that respects the network structure.

Proof This is trivial as for the perturbation to respect the network structure we require that the perturbed system be G admissible and it therefore satisfies the same relations.

Remark 1 Any asynchronous scheme on a graph $G = \langle V, E \rangle$ as outlined in the introduction gives rise to a G admissible IFS. We have a set of functions $\{f_v :$

$v \in V\}$ and we can also index the co-ordinates of the state space X with the elements of V such that $f_v(x)_u = x_v$ whenever u and v are non adjacent in G and $f_v(x)_v = F_v(x)$ and $\frac{\delta F_v(x)}{\delta x_u} = 0$ whenever u and v are non adjacent in G. Therefore if u and v are non adjacent we have

$$f_u \circ f_v(x_1, ..., x_v, ..., x_u, ..., x_d) = f_u(x_1, ..., F_v(x_w : (w, v) \in E), ..., x_u, ..., x_d) \tag{2.59}$$
$$= (x_1, ..., F_v(x_w : (w, v) \in E), ..., F_u(x_w : (w, u) \in E), ..., x_d)$$

and likewise for $f_v \circ f_u(x)$. So that an asynchronous system on G gives rise to a G admissible IFS.

Remark 2 Therefore in the context of networked systems commutivity relations are typical. If we have some physical system admitting a network structure then slight changes to the exact behavior of the agents playing out the dynamics will not affect the network structure and therefore leave the commutivity relations unchanged. We can therefore exploit these relations to make robust statements about the system.

For our example this means that some sequences of updates give rise to exactly the same behavior which gives rise to an equivalence relation on the update sequences.

$$\sigma|_n \bowtie \varsigma|_n \Leftrightarrow A(\sigma|_n) = A(\varsigma|_n) \tag{2.60}$$

We can now count D_n the number of equivalence classes of sequences of length n.

Theorem The exponent

$$\theta = \lim_{n\to\infty} \frac{\log D_n}{n} \tag{2.61}$$

is less than or equal to $\mathcal{X}$ the greatest eigenvalue of the matrix $\widehat{A}$ obtained by setting all diagonal and upper triangular elements of $\mathcal{G}$'s adjacency matrix $A(\mathcal{G})$ to one.

Proof Suppose that we have a representative σ for each $\bowtie|_n$ equivalence class. We will define a total order on the symbols in V then apply a sort of restricted

bubble sort algorithm to the truncated sequence $\sigma|_n$ to obtain a unique 'locally as increasing as possible' representative ς. The algorithm is as follows

1. Examine the pairs (σ_i, σ_{i+1}) in turn until finding a pair corresponding to non adjacent vertices with $\sigma_i > \sigma_{i+1}$
2. Swap this pair so that the sequence now reads $\sigma_1, ..., \sigma_{i-1}, \sigma_{i+1}, \sigma_i\sigma_{i+2}, ..., \sigma_n$
3. Continue until no pair satisfies the conditions required for a swap.

Claim The restricted bubble sort algorithm terminates

Proof The algorithm gives rise to a deterministic finite state space dynamical system and all orbits are therefore eventually periodic. Suppose that the algorithm does not terminate, then there is a periodic orbit to the algorithm. Now suppose that $i \neq j$ are two symbols that appear in the some periodic sequence and that during the evolution of the algorithm at least one (i, j) or (j, i) swap occurs. By periodicity at least one (j, i) or (i, j) swap must also occur in order for us to return to the original state and since either $i > j$ or $j > i$ one of these swaps is invalid and a contradiction.

Thus each equivalence class can be identified with a locally increasing sequence and $D_n \leq \widehat{D}_n$ where $\widehat{D}_n$ is the number of locally increasing sequences of length n. The set of all locally increasing sequences of length n is given by $\widehat{I} = \{\sigma|_n s.t. \sigma(k) = i, \sigma(k+1) = j \Rightarrow \text{either } i < j \text{ or } i \text{ and } j \text{ are connected in } G\}$ or equivalently

$$\widehat{I} = \{\sigma|_n : \sigma(k) = i, \sigma(k+1) = j \Rightarrow \widehat{A}_{i,j} = 1\} \tag{2.62}$$

Where $\widehat{A}$ is the matrix obtained by setting all diagonal and upper triangular elements of $\mathcal{G}$'s adjacency matrix to one.

Claim $\widehat{D}_n = \|l(n)\|_1$ where $l(n)_i$ is the number of length n locally increasing sequences ending in the symbol i and $l(n+1) = \widehat{A}l(n)$

Proof It is easy to see that $l(n+1) = \widehat{A}l(n)$ since the number of length $n+1$ locally increasing sequences ending in the symbol i is the number of length n locally

increasing sequences that are still locally increasing if we add the symbol i to their end.

Now $\widehat{D}_n = \|\widehat{A}^n l(0)\|_1$ and since $\widehat{A}$ is irreducible and positive it has a unique greatest eigenvalue $\mathcal{X}$ which is positive and associated with a positive eigenvector v with nonzero weight on all co-ordinates so that

$$\lim_{n\to\infty} \frac{\log \widehat{D}_n}{n} = \lim_{n\to\infty} \frac{\log \mathcal{X}^n \langle l(0), v\rangle}{n} = \mathcal{X} \geq \theta \tag{2.63}$$

We are now ready for the main theorem

Theorem For a network admissible IFS

$$\dim_{BOX}(\nu) \leq \frac{\log \theta}{-\lambda} \leq \frac{\log \mathcal{X}}{-\lambda} \tag{2.64}$$

Proof Consider the sequence of sets

$$C_n = \bigcup_{\sigma|_n \in S|_n} \phi(\sigma|_n)(\Delta_d) \tag{2.65}$$

where the union is taken over all length n sequences. Since composition is taken backwards

$$\phi(\sigma|_{n+1})(\Delta_d) \subset \phi(\sigma|_n)(\Delta_d) \tag{2.66}$$

then $C_{n+1} \subset C_n$ so $\lim_{m\to\infty} C_m \subset C_n$ and $\nu(C_n) = 1$. Now many of the sequences are equivalent so we can instead take the union over a set of equivalence classes.

$$C_n = \bigcup_{\sigma_n \in \frac{S_n}{\bowtie|_n}} \phi(\sigma|_n)(\Delta_d) \tag{2.67}$$

For a μ measure one set of sequences $l(\sigma)|_n$ the diameter of the set $\phi(\sigma|_n)(\Delta_d)$ satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log l(\sigma)|_n = \lambda \tag{2.68}$$

We can therefore represent a measure one set of equivalence classes with a sequence satisfying the above limit. Thus C_n is union of $\approx \theta^n$ sets of diameter $\approx e^{\lambda n}$ and

$$\dim_{BOX}(\nu) \leq \frac{\log \theta}{-\lambda} \tag{2.69}$$

The second inequality follows from the previous theorem.

Remark This result therefore gives us a way of characterizing the measure associated with a network admissible IFS but moreover gives us a way of characterizing different network topologies. For a fixed Lyupanov exponent (which is admittedly a bit of a fiddle) a network with a smaller eigenvalue $\mathcal{X}$ will tend to induce a more singular measure. We might use the following rule of thumb; less connectivity $\Rightarrow$ more commutivity $\Rightarrow$ smaller $\mathcal{X} \Rightarrow$ more singular measure. Therefore $\mathcal{X}$ is a natural way to characterize the degree of connectivity in a network when we are considering an asynchronous process evolving on it.

2.3.2 Example

We can now obtain the bound for any example we like by calculating $\mathcal{X}$ then using our upper bound for λ from 1.3. We choose the following illustrative example as the two different networks have the same λ upper bound and therefore make for a 'fair' comparison.

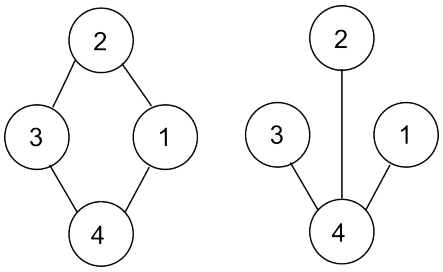


Figure 2.6: Example networks, ring and tree.

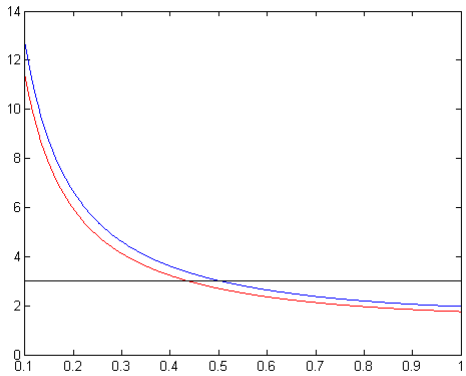


Figure 2.7: Measure dimension upper bound vs epsilon for the two systems on four vertices. Ring (blue) and tree (red). ν is guaranteed to be singular when the dimension is less than three, the dimension of Δ_4.

For the system on the ring we have $\theta = 3.4142$ and for the tree $\theta = 3$. So that at the same lyupanov exponent the system on the line will have a more singular measure. Of course the value of the exponent will usually also depend on the network in a complicated way so we should be careful in making any direct comparisons. Although in this case the figure clearly shows that the measure associated with the system on the tree is more singular.

2.4 Conclusion

We have shown that our class of linear systems converge to a random stable equilibrium. We have shown that an IFS formulation can be explicitly obtained which enables us to better examine the distribution on equilibrium values. In particular by counting the equivalence classes of update sequences we bound the dimension of the induced measure from above. This in tern gives us a neat way to characterize different networks with the eigenvalue $\mathcal{X}$ which roughly tells us how singular we expect the induced measures to be.

The IFS formulation relies on us being able to take the adjoint of our update functions and they must therefore be linear for this approach to work. The results of

section 3 will be perfectly valid for a non-linear system provided it is already in an COAIFS form which our original system was not. For instance the set of equilibrium points is a non-trivial fix subspace so the system is not a contraction. Of course it would be possible to linearize a non-linear system in the vicinity of an equilibrium to obtain a system with an COAIFS formulation and the eigenvalue $\mathcal{X}$ may still be useful for comparing different networks even when we are only interested in non-linear dynamics.

Bibliography

[1] Gilbert N. (2007). Agent-Based Models *SAGE Publications*

[2] Malinchik S. (2010). Framework for Modeling Opinion Dynamics Influenced by Targeted Messages *SoicalCom* Pages 697-700

[3] Bracha G, Toueg S. (1985). Asynchronous Consensus and Broadcast Protocols, *JACM*, Volume 32, Pages 824-840

[4] Diaconis P., Freedman D. (1999). Iterated random functions, *SIAM Review*, Volume 41, Pages 41-76

[5] Falkoner K. (2003). Fractal Geometry: Mathematical Foundations and Applications *Wiley*

[6] Nicol M, Sidirov N, Broomhead D. (2001) On the Fine Structure of Stationary Measures in Systems which Contract-on-Average *Journal of Theoretical Probability*, Volume 15, Pages 715-739

Chapter 3

Iterative solution of graph Laplacians: Synchrony vs Asynchrony

Abstract

Iterative relaxation is the numerical technique used to solve the equation $Ax = b$ by simulating the dynamical system $x(n+1) = (\epsilon A - I)x(n) - \epsilon b$. As well as reviewing the standard implementation of this method we consider an asynchronous implementation where the individual nodes (co-ordinates of x) update themselves independently of the others in a random order. This procedure can in principal be carried out much faster, especially on an asynchronous CPU or over a network of parallel CPUs. Further computational advantage can be obtained if A is a sparse matrix, where we can identify the support of A with a graph of relatively few connections. Focusing on graph Laplacians we show that the asynchronous relaxation converges to a solution of the original problem and obtain a sandwich inequality for the rate of convergence.

3.1 Introduction

The Laplacian of a graph G on $\{1, 2, ..., N\}$ is an $N \times N$ matrix L with

$$L_{i,j} = \begin{cases} -1 & \text{if } i = j \\ \frac{1}{d_i} & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

where d_i is the in degree of i in G. In analogy to the Laplacian operator ∇^2, L can be used to describe diffusion processes on a graph, in particular the dynamical system

$$x(n+1) = (I - \epsilon L)x(n) \tag{3.2}$$

corresponds exactly to the standard spatiotemporal discretization of the heat equation when G is taken from a regular grid partitioning the spatial domain.

Boundary conditions can be added to these systems by forcing some subset of vertices to maintain fixed boundary values, again in analogy to the PDE case. If we have a graph G on $N + M$ vertices where the last M are all boundary vertices then the state of the non-boundary vertices evolves according to

$$x(n+1) = (I - \epsilon L|_{G/B})x(n) + \epsilon b \tag{3.3}$$

where $L|_{G/B}$ is the graph Laplacian restricted to G/B defined by

$$L_{i,j} = \begin{cases} -1 & \text{if } i = j \\ \frac{1}{d_i} & \text{if } (i,j) \text{ is an edge in } G/B \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

but d_i is still the in degree of i in G, not in G/B. And b - as in (3.3) - is defined by

$$b_i = \frac{1}{d_i} \sum_{j \in B_i} x_j \tag{3.5}$$

where B_i is a subset of the boundary containing vertices with directed edges to i and the x_j are the corresponding boundary values.

This chapter is concerned with finding solutions to $L|_{G/B}x = b$ by iterating (3.3). In particular we are interested in the rate of convergence of this procedure when (3.3) is implemented in an asynchronous way.

3.1.1 Asynchronous iteration

Suppose that instead of updating all of the co-ordinates of x simultaneously we assigned a separate processor to each vertex of G which was able to broadcast its state down all of its outwards edges and receive the current states of its neighbors through all of its inwards edges. The individual processers could then independently compute the next value of their vertices state and update it as soon as the computation was finished, rather than having to wait until all the other vertices where ready. In principal the average 'update' rate per vertex of this asynchronous system could be faster than the simultaneous version as there would be less time spent waiting for other slower processes to finish.

A fully realistic model of the sequence of inter-update times for a particular vertex is beyond the scope of this work, instead we use a random sequence with some appealing mathematical properties. We assume that each vertex updates its state independently of all the others with i.i.d. mean-ν exponentially distributed inter-update times. The main advantage of this formulation is that the sequence of updates, that is a sequence in the vertex set given by the order they update in, is Bernoulli distributed. The *Asynchronous iteration* of (3.3) is then given by

$$x(\sigma, n) = A(\sigma_n)x(\sigma, n-1) + b(\sigma_n) \tag{3.6}$$

where $A(i)$ is the $N \times N$ identity except for the ith row which is taken from $I - \epsilon L|_{G/B}$, $b(i) = e_i b_i$ and σ is a $\{1, 2, ..., N\}^{\mathbb{N}}$-valued random variable with uniform bernoulli distribution.

This discrete time system is reconcilable with the continuous time system outlined through the law of large numbers, $T(n)$ the time of the nth update satisfies (with probability one)

$$\lim_{n\to\infty} \frac{T(n)}{n} = \frac{\nu}{N} \tag{3.7}$$

3.2 Synchronous system

The synchronous iteration is given by

$$x(n) = (I - \epsilon L|_{G/B})x(n-1) + \epsilon b = Ax(n-1) + b \tag{3.8}$$

In what follows we will require that G/B is irreducible, that there is a directed path between any two vertices, and aperiodic, that the greatest common divider of the cycles in G/B is one. These assumptions are required for applying standard results to the matrix A, they are not entirely necessary though, for instance G/B may be comprised of two disconnected components which could be thought of as two separate irreducible systems. Further degeneracies can be treated similarly.

Theorem The synchronous system converges exponentially quickly to the unique equilibrium $x = L|_{G/B}^{-1}b$ with exponent $\log\lambda$, where λ is the unique maximal eigenvalue of the matrix A.

Proof Since our system is linear we can separate the equilibrium from the state so that

$$x(n) = L|_{G/B}^{-1}b + [x(n) - L|_{G/B}^{-1}b] \tag{3.9}$$

and

$$\begin{aligned} x(n+1) &= (I - \epsilon L|_{G/B})L|_{G/B}^{-1}b + A[x(n) - L|_{G/B}^{-1}b] + \epsilon b \\ &= L|_{G/B}^{-1}b + A[x(n) - L|_{G/B}^{-1}b] = L|_{G/B}^{-1}b + \underbrace{A^n[x(0) - L|_{G/B}^{-1}b]} \end{aligned} \tag{3.10}$$

So the difference between the equilibrium and the state of our system at stage n is given by the underbraced term. The matrix A is irreducible and aperiodic so has a simple maximal eigenvalue $0 < \lambda < 1$ with an eigenvector u which is strictly positive. Therefore this error term will decay like λ^n.

3.3 Asynchronous system

As in the synchronous case we are able to separate the equilibrium from the state so that

$$x(\sigma, n) = L|_{G/B}^{-1}b + [x(\sigma, n) - L|_{G/B}^{-1}b] \tag{3.11}$$

and

$$x(\sigma, n+1) = A(\sigma_{n+1})L|_{G/B}^{-1}b + A(\sigma_{n+1})[x(n) - L|_{G/B}^{-1}b] + \epsilon b(i) \tag{3.12}$$

$$= L|_{G/B}^{-1}b + A(\sigma_{n+1})[x(\sigma, n) - L|_{G/B}^{-1}b] = L|_{G/B}^{-1}b + \underbrace{[\prod_{k=1}^{n+1} A(\sigma_k)][x(\sigma, 0) - L|_{G/B}^{-1}b]}$$

So that once again our state is given by the equilibrium plus an error which evolves independently to the solution. The stability of this system is determined by the exponent

$$\mu(\sigma) = \lim_{n\to\infty} \frac{1}{n} \sup_x \log \frac{||\prod_{k=1}^{n} A(\sigma_k)x||}{||x||} \tag{3.13}$$

whose existence and uniqueness for a measure one set of sequences, σ is guaranteed by the multiplicative ergodic theorem so we are justified in simply writing μ. In order to show that the asynchronous system converges to the equilibrium we must show that $\mu < 0$.

3.3.1 Bounds on μ

We will exploit some strong non-negativity properties to prove both of our bounds on μ.

Theorem We have the following bounds

$$\log \lambda \leq N\mu \leq N \log(1 + \frac{\lambda - 1}{N}) \tag{3.14}$$

The proof follows the supporting lemma below

Lemma The exponent can be re-expressed by

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sup_{x'} \log \frac{||\prod_{k=1}^{n} A(\sigma_k)x'||}{||x'||} \tag{3.15}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \langle \prod_{k=1}^{n} A(\sigma_k)x, u \rangle$$

For any strictly-positive x. Recall that u is the eigenvector corresponding to A's unique maximal eigenvalue.

Proof With probability one

$$P(\sigma, n) = \prod_{k=1}^{n} A(\sigma_k) \tag{3.16}$$

is strictly positive for all n greater than some finite n_0. We must prove this componentwise, for $i \neq j$ the (i,j)th component is non-decreasing and will increase strictly on the application of $A(\varsigma_M = i)A(\varsigma_{M-1})...A(\varsigma_1 = j)$ where ς is the shortest path from j to i, with probability one we will eventually apply this sequence of updates. The (i,i)th component is only affected by the ith vertex updating when its new value is bounded below by $(1 - \epsilon d_i)$ times the old value, this remains positive so our claim is true.

The matrix $P(\sigma, n)$ is therefore irreducible and aperiodic so has a maximal simple eigenvalue $\gamma(\sigma, n)$ corresponding to a strictly positive eigenvector $v(\sigma, n)$. The supremum in (3.15) is therefore attained by this eigenvector and

$$\mu = \lim_{n\to\infty} \frac{1}{n} \log \gamma(\sigma, n) \tag{3.17}$$

The second expression in (3.15) becomes

$$\lim_{n\to\infty} \frac{1}{n} \log\langle \gamma(\sigma, n) \underbrace{v(\sigma, n) \min_j \frac{x_j}{v_j}} + \prod_{k=1}^{n} A(\sigma_k) y, u\rangle \tag{3.18}$$

where y is the difference between the non-zero underbraced term and x. Now since y is non-negative we have

$$\geq \lim_{n\to\infty} \frac{1}{n} \log \gamma(\sigma, n) + \frac{1}{n} \log\langle v(\sigma, n) \min_j \frac{x_j}{v_j}, u\rangle \geq \mu \tag{3.19}$$

And since our original expression for μ was a maximum which gives the opposite non-strict inequity our two expressions are the same.

Upper bound $\mu < \log(1 + \frac{\lambda - 1}{N}) < 0$

Proof Since μ is a constant for almost all sequences we can express it as an expectation

$$\mu = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_\sigma \log\langle \prod_{k=1}^{n} A(\sigma_k) x, u\rangle \tag{3.20}$$

and the expectation of a log is less than the log of an expectation so

$$\mu < \lim_{n\to\infty} \frac{1}{n} \log \underbrace{\mathbb{E}_\sigma \langle \prod_{k=1}^{n} A(\sigma_k)x, u\rangle} \tag{3.21}$$

By linearity of expectation we can take the expectation inside the inner product.

$$\mathbb{E}_\sigma \prod_{k=1}^{n} A(\sigma_k) = \frac{1}{N^n} \sum_{\sigma|_n} \prod_{k=1}^{n} A(\sigma_k) = \frac{1}{N^{n-1}} \sum_{\sigma|_{n-1}} \sum_{i=1}^{N} \frac{1}{N} A(i) \prod_{k=1}^{n-1} A(\sigma_k) \tag{3.22}$$

$$= [\sum_{i=1}^{N} \frac{1}{N} A(i)] \frac{1}{N^{n-1}} \sum_{\sigma|_{n-1}} \prod_{k=1}^{n-1} A(\sigma_k) = [\sum_{i=1}^{N} \frac{1}{N} A(i)]^n = (I + \frac{A-I}{N})^n$$

so that

$$\mu < \lim_{n\to\infty} \frac{1}{n} \log\langle (I + \frac{A-I}{N})^n x, u\rangle = \lim_{n\to\infty} \frac{1}{n} \log\langle x, u(I + \frac{A-I}{N})^n\rangle \tag{3.23}$$

$$= \log(1 + \frac{\lambda - 1}{N}) + \lim_{n\to\infty} \frac{1}{n} \log\langle x, u\rangle = \log(1 + \frac{\lambda - 1}{N})$$

Lower bound $\mu \geq \frac{1}{N} \log \lambda$

Proof Again we use the expectation expression for μ

$$\mu = \lim_{n\to\infty} \frac{1}{n} \frac{1}{N^n} \sum_{\sigma|_n} \log\langle \prod_{k=1}^{n} A(\sigma_k)x, u\rangle \tag{3.24}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \frac{1}{N^m} \underbrace{[\sum_{\sigma|_m} \log\langle \prod_{k=1}^{m} A(\sigma_k)x, u\rangle - N \sum_{\sigma|_{m-1}} \log\langle \prod_{k=1}^{m-1} A(\sigma_k)x, u\rangle]}$$

and the underbraced term can be rewritten as

$$\sum_{\sigma|_{m-1}} [\sum_{i=1}^{N} \log\langle A(i) \prod_{k=1}^{m-1} A(\sigma_k)x, u\rangle] - N \log\langle \prod_{k=1}^{m-1} A(\sigma_k)x, u\rangle \tag{3.25}$$

$$= \sum_{\sigma_{m-1}} \overbrace{\log \prod_{i=1}^{N} \frac{\langle A(i)y, u\rangle}{\langle y, u\rangle}}$$

where $y = \prod_{k=1}^{m-1} A(\sigma_k)x$. Using $A(i) = I + [A(i) - I]$ the overbraced term is given by

$$\log \prod_{i=1}^{N} \frac{\langle y, u\rangle + \langle y, u[A(i) - I]\rangle}{\langle y, u\rangle} \tag{3.26}$$

and since $A(i)$ only acts on u's ith component we have

$$= \log \prod_{i=1}^{N} [1 + \frac{\langle y, (\lambda - 1) u_i e_i \rangle}{\langle y, u \rangle}] = \log \prod_{i=1}^{N} [1 + \frac{(\lambda - 1) u_i y_i}{\langle y, u \rangle}] \tag{3.27}$$

$$= \log[1 + (\lambda - 1) \sum_{i=1}^{N} \frac{u_i y_i}{\langle y, u \rangle} + (\lambda - 1)^2 \sum_{i \neq j} \frac{u_i y_i u_j y_j}{\langle y, u \rangle^2} + ...] = \log[\lambda + g(y)]$$

where

$$g(y) = \sum_{k=1}^{N} (-1)^k (1 - \lambda)^k \frac{\sum_{i_1 \neq i_2 \neq ... \neq i_k} \prod_{j=1}^{k} u_{i_j} y_{i_j}}{\langle y, u \rangle^k} \tag{3.28}$$

$$\geq \sum_{k=1}^{N/2} \frac{(1 - \lambda)^{2k}}{\langle y, u \rangle^{2k}} [\sum_{i_1 \neq i_2 \neq ... \neq i_{2k}} \prod_{j=1}^{k} u_{i_j} y_{i_j} - \frac{(1 - \lambda)}{\langle y, u \rangle} \sum_{i_1 \neq i_2 \neq ... \neq i_{2k+1}} \prod_{j=1}^{k} u_{i_j} y_{i_j}] \tag{3.29}$$

$$= \sum_{k=1}^{N/2} \frac{(1 - \lambda)^{2k}}{\langle y, u \rangle^{2k}} \sum_{i_1 \neq i_2 \neq ... \neq i_{2k}} [\prod_{j=1}^{k} u_{i_j} y_{i_j} \underbrace{(1 - (1 - \lambda) \sum_{t \neq i_1, ..., i_{2k}} \frac{u_t y_t}{\langle y, u \rangle})}]$$

So since

$$\sum_{t \neq i_1, ..., i_{2k}} \frac{u_t y_t}{\langle y, u \rangle} \leq 1 \tag{3.30}$$

and $(1 - \lambda) < 1$ the underbraced term in (3.29) is positive and $g(y)$ is positive for all y. We can therefore replace equality with $\log[\lambda + g(y)]$ in (3.27) with greater than or equal to $\log \lambda$. Substituting back into (3.24) we have

$$\mu \geq \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \frac{1}{N^m} \sum_{\sigma|_{m-1}} \log \lambda = \frac{1}{N} \log \lambda \tag{3.31}$$

3.4 Examples

We simulate a graph laplacians for lines of vertices with boundary nodes at the ends of the lines. To ensure that our approximation of λ is accurate we repeatedly renormalize the state variable and use the series of renormalization factors to compute the exponent.

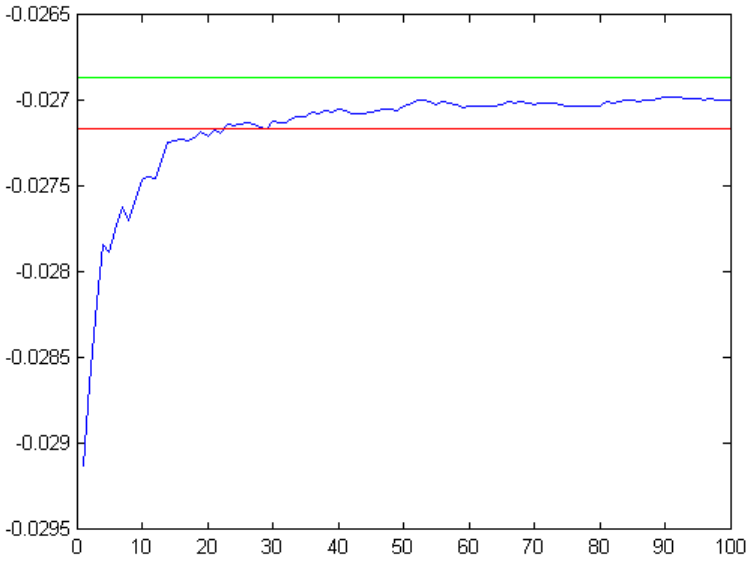


Figure 3.1: Convergence of μ, for a line with $N = 5$ $\epsilon = 0.1$ n in 1000s

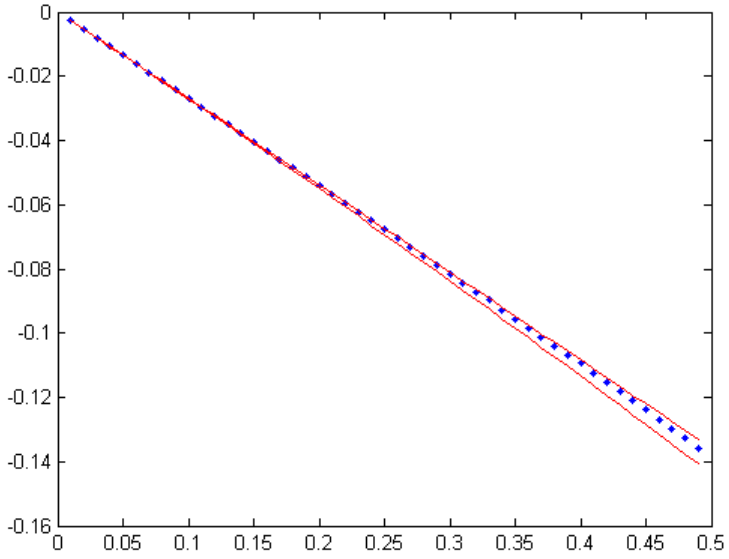


Figure 3.2: μ and bounds vs ϵ for $N = 5$

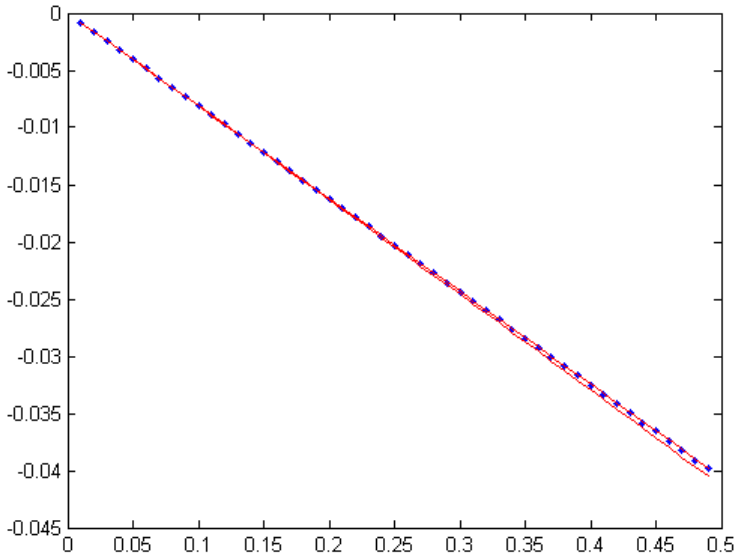


Figure 3.3: μ and bounds vs ϵ for $N = 15$

3.4.1 Remarks on asynchronous computing: Synchrony Vs Asynchrony

Comparing the rate of convergence in our two systems we have two different time scales; ν_S the average interupdate time of the synchronous system and $\frac{\nu_A}{N}$ the average interupdate time of the asynchronous system where ν_A is the average interupdate time of an individual node in the asynchronous system. Our bounds show that if $V_S = V_A$ then we have roughly the same rate of convergence with the asynchronous system very slightly slower. However if it typically takes a single processor C seconds to perform the computation associated with a single node updating then a single processor implementing the synchronous scheme has $\nu_S = NC$ whereas N parallel processors implementing the asynchronous scheme have $\nu_A = C$ so that the asynchronous system is roughly N times faster.

3.4.2 Conclusion

Of course our model does not serve as a physically realistic description of any parallel computation used in practice. There is no accounting for latency in communication between the processors and the exponential interupdate times where chosen for mathematical convenience and have no real justification.

It is clear from experimental computer science that under the right circumstances real life asynchronous parallel schemes can vastly outperform traditional synchronous approaches, however the exact detail of the implementation is extremely important [1, 2, 3].

In [1] Dingle et al employ an asynchronous scheme to solve $Ax = b$ using Jacobi's method

$$x_i(n) = \frac{1}{a_{i,i}}(b_i - \sum_{j \neq i} a_{i,j} x_j(n-1)) \tag{3.32}$$

by evaluation $x_i(n)$ in parallel for each i. In future work we hope to instrument this procedure to provide us with statistical information about the update sequences which we can use to develop a new more realistic, but still tractable model of the process.

By using a bernoulli update sequences the work in this example serves as a first approximation to this more complex model.

Bibliography

[1] Iain Bethune, J. Mark Bull, Nicholas J. Dingle and Nicholas J. Higham. (2011). Investigating the Performance of Asynchronous Jacobi's Method for Solving Systems of Linear Equations *mims eprint*

[2] D. Chazan, W. Miranker. (1968). Chaotic relaxations *Linear Algebra and its Applications* Volume 2, Issue 2, Pages 199-222

[3] Douglas de Jager, Jeremy T. Bradley (2010). Extracting State-Based Performance Metrics using Asynchronous Iterative Techniques*Performance Evaluation.* Volume 67, Issue 12, Pages 1353 - 1372

Chapter 4

Stochastic production trees as products of i.i.d. fixed support componentwise exponential max-plus matrices

Abstract

We introduce a class of stochastic production tree model, based on Petri nets, which admit a random matrix product description in the Max-plus algebra. With a kind of combinatorial change of variables we are able to simplify the form of the matrices arising from these models. For this class of *Componentwise exponential* matrix, whose non-zero components are exponentially distributed, we prove a new result relating the (Max-plus) spectrum of the product to the principal (classical) eigenvalue of an associated adjacency matrix by means of a sandwich inequality. This theorem highlights several important theoretical factors in the dynamics of Max-plus linear systems generally and gives us some neat insight into the different production tree models.

Introduction

The Max-plus algebra gives us an alternative way to look at a number of interesting classically non-linear phenomena [1, 2, 3, 4, 5, 6]. In particular there are many quite different dynamical systems/mathematical constructions which are quite intractable with standard algebra but become linear or in some other way simpler when expressed in Max-plus.

Dynamical systems whose variables are the starting times of different interacting events fit quite naturally into the Max-plus linear systems framework. If some event i can only reoccur for the $(n+1)$th time $M_{i,j}$ seconds after event j has occurred for the nth time then $x_i(n+1)$, the time i occurs for the $(n+1)$th time, satisfies

$$x_i(n+1) = \max_j M_{i,j} + x_j(n) = [M \otimes \underline{x}(n)]_i \tag{4.1}$$

where $\otimes$ stands for Max-plus multiplication - which for elements of $\mathbb{R} \cup -\infty$ is simply addition, for a matrix vector pair is as defined in (4.1) and for a pair of max-plus matrices $[A \otimes B]_{i,j} = \max_k A_{i,k} + B_{k,j}$. The theory for Max-plus matrix algebra follows in analogy to the classical case; there are eigenvalues and eigenvectors, determinants, the Caley-Hamilton theorem holds and much more besides [6]. For an introduction to Max-plus algebra see [1].

This chapter focuses on stochastic queuing systems which can also be thought of as very high dimensional continuous time Markov chains. Although they are classically linear their dimension makes them intractable and investigating any classical structure in these models seems to be beyond the scope of standard Markov chain methods. We are however to describe them completely with much lower dimensional Max-plus linear models whose underlying structure can be studied more easily.

We show that products of i.i.d. Max-plus matrices are dominated by a Max-plus exponent whose value can be attributed to the weight of a path through the vertex set of the matrices. By taking this path-centric viewpoint we are able to give a new neater proof of the Max-plus multiplicative ergodic theorem and prove a new result relating the Max-plus exponent to the classical principal eigenvalue of an associated adjacency matrix.

In section 1 we introduce petri-nets and explain how stochastically timed event graphs can be described by products of i.i.d. max-plus matrices. In 1.1 we introduce the main production line example. In section 2 we prove the max-plus multiplicative ergodic theorem then introduce componentwise exponential max-plus matrices, which are arguably the simplest possible non-trivial random max-plus matrices with arbitrary graph topology, and prove my main result; a sandwich of bounds for the max-plus exponent based on the principal eigenvalue of an associated graph. In section 3 we generalize the first example to a class of production trees. We show how these systems can be recast in a componentwise exponential form and then examine their exponents. Our main theorem allows us to rigourously investigate the asymptotic behavior of the production tree systems.

4.1 Stochastic timed event graphs

Petri-nets are a modeling language that can be used to describe a wide variety of distributed systems [7]. Formally a petri-net is a directed bipartite graph in which the vertices are either *transitions* (signified by bars, which represent some events that can occur) or *places* (signified by circles, which represent some conditions or processes). Places can contain one or more *tokens* which should be thought of as units of information or of some product. Each token can be in a waiting state (signified by a grey dot in the place) or ready state (signified by a black dot in the place). Waiting tokens become ready after some *waiting-time* which is particular to the place that the token occupies and can be a fixed non-negative number or random variable. The dynamics of the petri-net move tokens from place to place.

Each transition that succeeds some place-vertices can become *enabled*. A transition is enabled if and only if each of its predecessors contains at least one ready token. When a transition is enabled it is able to fire. When a transition fires a single ready token is removed from each place it succeeds and a waiting token is added to each place it is a predecessor of.

Note that since it is possible to have two enabled transitions when firings are

mutually exclusive, the evolution of a petri net is not necessarily deterministic even if the place waiting-times are constant. See Fig 1.

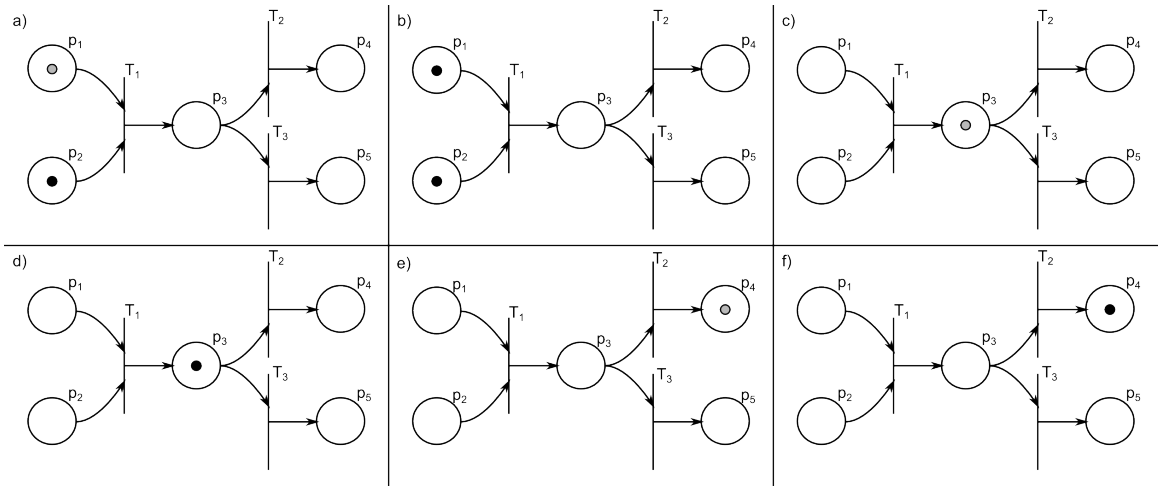


Figure 4.1: Evolution of a simple petri-net. a) The token in p_1 must wait some time to become ready, b) Now that there are ready tokens in p_1 and p_2 the transition T_1 is enabled, c) On firing T_1 takes the ready tokens from p_1 and p_2 and adds a waiting token to p_3, d) Now that the token in p_3 is ready both transitions T_2 and T_3 are enabled, e) In this example we choose T_2 to fire taking the token from p_3 and adding a waiting token to p_4, f) The token in p_4 is now ready.

Definition A *stochastic timed event graph* is a petri net with the following properties

- Each place has exactly one predecessor and one successor transition.
- The successive waiting times at a place i are an i.i.d. sequence $[t_i(n)]_{n=1}^{\infty}$ independent of the waiting times at all other places.

The first condition guarantees that the only non-determinism in the system comes from the random waiting times and not from some 'higher' controller as in the example in Fig 1. The second condition enables us to describe the evolution of the dynamics with a product of i.i.d max-plus matrices.

4.1.1 Example - Asynchronous production line

The following stochastic timed event graph can be used to model a fairly general asynchronous production line. Consider a plant where we have a production line consisting of N sites where each item being produced must pass from one site to the next undergoing a different process at each site, each taking a random period of time. Only one item may occupy a site at a time so that if one site is taking a long time to process an item a queue may build up behind it.

Therefore whenever an item arrives at a site the processing begins. When it is processed the item is either instantaneously moved for processing at the next site, or if the next site is occupied, must wait until the next site is unoccupied when it is instantaneously moved on.

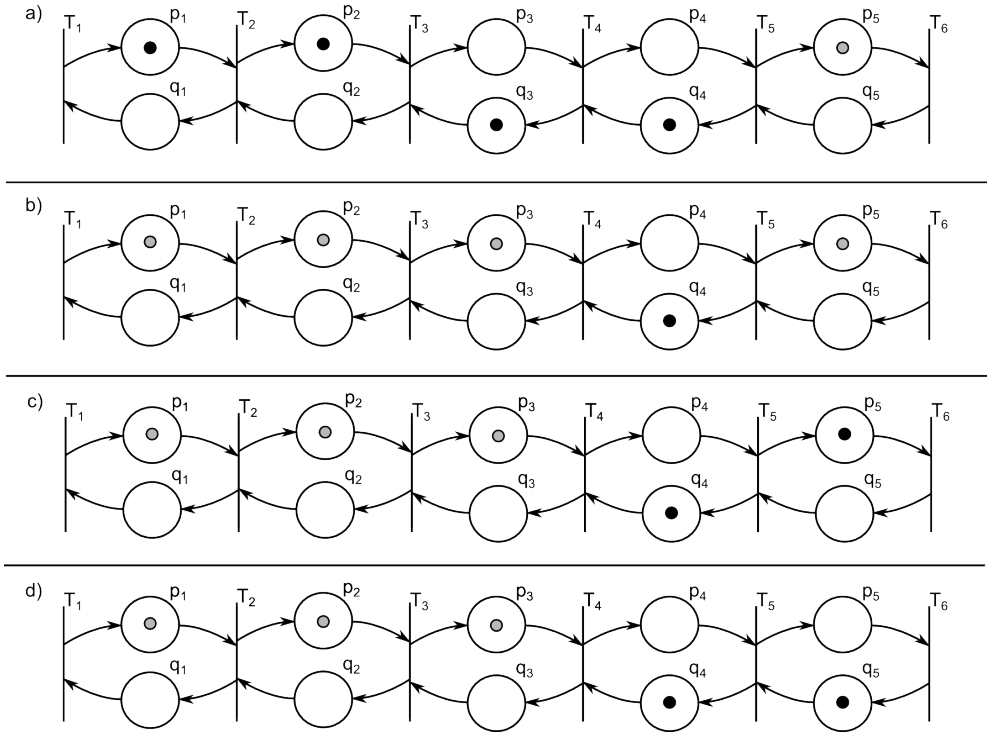


Figure 4.2: Evolution of asynchronous production line for $N = 5$. a) Transition T_3 is enabled, b) Transition T_3 fires instantaneously enabling T_2 which fires instantaneously enabling T_1 which also fires, c) Transition T_6 is enabled, d) Transition T_6 fires.

To model this queue with a timed event graph we divide the places up into two

groups $\{p_i\}$ and $\{q_i\}$ with $i = 1, 2, ..., N$. See fig 2. The $\{p_i\}$ represent the sites in the production line so that a waiting token in place p_i should be thought of as an item undergoing a process at site i, likewise a ready token at place i should be thought of as a processed item at site i. The waiting times at p_i are a sequence of i.i.d. random variables $[t_i(n)]_{n=1}^{\infty}$ with distribution given by the time that the ith process takes.

The $\{q_i\}$ are used to make sure that only one token occupies each p_i at a time, a token in place q_i indicates that the ith site being empty and ready to receive a new item for processing. The waiting times at q_i are all zero so that all tokens in $\{q_i\}$ are instantaneously ready.

Transitions T_i with $i = 2, 3, ..., N-1$ represent a processed item a site $i-1$ being transferred to site i. The initiating transition T_1 represents a new item being brought into the line and the terminal transition T_N represents the completion of a finished product.

This system could be used to model all sorts of different production lines, most interestingly it gives a very complete description of a bio-molecular process called mRNA transcription, the process by which ribosomes build proteins coded for in mRNA. For a thorough analysis of the deterministically timed version of the model for this application see [8].

4.1.2 Max-plus formulation

Given an initial distribution of tokens (which we call a marking) in a stochastic timed event graph the evolution of the petri-net's state can be completely determined by the sequence of transition firing times. This statement is trivial to prove, suppose that we start with some initial marking m_0 and arrive after some time at a new marking m'. If we are given the transition firing sequence but not the sequence of states that the system moves through we can simply apply these transitions to m_0 until we arrive at m'.

The state variable for our max-plus model at stage n is then given by the firing time vector $X(n) = [x_i(n)]$ where $x_i(n)$ is the time at which transition i fires for the

n'th time.

All stochastic timed event graphs satisfy a max-plus linear equation which can easily be obtained [1, 2]. Since all the systems we will consider have the special property that each place only ever hold zero or one tokens the formulation is slightly simpler. At each stage the firing time vectors satisfy

$$X(n) = A(n) \otimes X(n) \oplus B(n-1) \otimes X(n-1) \tag{4.2}$$

Where $A(n)$ and $B(n-1)$ are max-plus matrices whose components are determined by the random waiting times associated with the places in the petri net and M_o the initial marking of tokens.

$$A(n)_{i,j} = \epsilon \oplus \bigoplus_{\{k:T_j \to p_k \to T_i, m_0(k)=0\}} t_k(n) \tag{4.3}$$

$$B(n)_{i,j} = \epsilon \oplus \bigoplus_{\{k:T_j \to p_k \to T_i, m_0(k)=1\}} t_k(n) \tag{4.4}$$

So that $A(n)_{i,j}$ is a random variable given by the maximum of a set of the nth place waiting times or equal to $-\infty$ if this set is empty. This maximum is taken over all the places that join T_j to T_i and contain no tokens in the initial marking. Likewise $B(n)_{i,j}$ is a maximum but taken over places that contain one token in the initial marking.

Since when we fix the random waiting times our system's evolution is deterministic, it follows that $X(n)$ exists and is unique. It can also be shown that (1) has a unique solution [6]. Define the Kleene star of $A(n)$ by

$$A(n)^* = I \oplus A(n) \oplus A(n)^{\otimes 2} \oplus A(n)^{\otimes 3} \oplus ... \tag{4.5}$$

So that $A(n)^*_{i,j}$ should be thought of as the weight of the maximally weighted path from j to i of any length through $A(n)$'s associated graph. Unless our petri-net's initial marking contains a circuit with no tokens in any of its places $A(n)^*$ will exist since $A^{\otimes K} = \epsilon$ for all $K > N$. Any petri-net/initial marking with such a circuit will have trivial dynamics since none of the transitions will be able to fire. We can use the Kleene star to construct a solution to (1) by

$$X(n) = A(n)^* \otimes B(n-1) \otimes X(n-1) \tag{4.6}$$

which on substitution to the RHS yields

$$A(n) \otimes A(n)^* \otimes B(n-1) \otimes X(n-1) \oplus B(n-1) \otimes X(n-1) \tag{4.7}$$

$$= [A(n) \otimes A(n)^* \oplus I] \otimes B(n-1) \otimes X(n-1) = A(n)^* \otimes B(n-1) \otimes X(n-1)$$

Therefore this is the unique solution to (1) and provides the unique evolution of our system. The dynamics of our stochastic timed event graph are governed by the max-plus linear system

$$X(n) = A(n)^* \otimes B(n-1) \otimes X(n-1) = [\bigotimes_{k=1}^{n} \underbrace{A(k)^* \otimes B(k-1)}] \otimes X(0) \tag{4.8}$$

And since the components of $A(k)$ and $B(k-1)$ are drawn from disjoint sets of edge weights the underbraced terms form a sequence of i.i.d. random max-plus matrices.

4.1.3 Example

We can now construct the max-plus linear system associated with our asynchronous production line model. We will use the initial making m_0 where there are no tokens in any of the p_i and one token in each q_i, this corresponds to starting the plant with no items currently in processing. As outlined we choose to represent max-plus matrices with their associated weighted graphs rather than as an array of numbers, see fig 3.

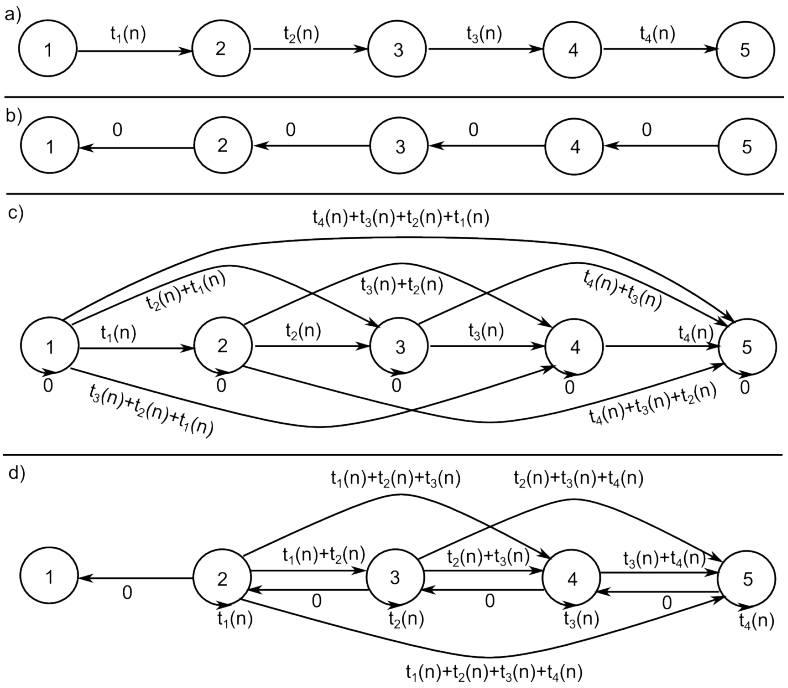


Figure 4.3: Graphs associated with $A(n)$,$B(n-1)$,$A(n)^*$ and $A(n)^*B(n-1)$ respectively for $N = 4$

We can now simulate our system by choosing an initial condition $X(0)$ then randomly generating a sequence of matrices $[A(k)^*B(k-1)]_{k=1}^n$ and max-plus multiplying the state with them to obtain $X(n)$, see section 2.1.

4.2 Max-plus Lyapunov exponent

Our max-plus linear systems are evolved according to

$$X(n) = M(n) \otimes X(n-1) = [\underbrace{\bigotimes_{k=1}^{n} M(k)}] \otimes X(0) \tag{4.9}$$

where the $[M(k)]_{k=1}^{\infty}$ are a sequence of i.i.d. random max-plus matrices. Since the associated graphs for these matrices have the same vertex and edge set each time, just different edge weights, we can interpret the underbraced matrix product which we denote $P(n)$ as follows. $P(n)_{i,j}$ is given by the weight of the maximally weighted path σ of length n through the associated graphs from j to i which accumulates

weight on its k'th step according to $M(k)$ so that σ's total weight is given by

$$W(\sigma) = \sum_{k=1}^{n} M(k)_{\sigma(k-1),\sigma(k)} \tag{4.10}$$

This maximally weighted path perspective is very natural to use in max-plus linear algebra and is essential to the new theory presented in this chapter. It also enables us to give a new, simpler proof of the max-plus multiplicative ergodic theorem.

Theorem Provided the graph associated with M is irreducible and aperiodic then with probability-1 the limit

$$\lambda = \lim_{n\to\infty} \frac{1}{n} P(n) = \lim_{n\to\infty} \frac{1}{n} \bigotimes_{k=1}^{n} M(k) \tag{4.11}$$

exists and is a matrix with each element equal to the same constant, the *max-plus Lyapunov exponent* which we shall also denote λ. Note that since conventional multiplication acts like taking powers in max plus we are justified in calling this an exponent.

Proof This proof is a slightly different to the standard treatment which follows from Kingman's sub-additive ergodic theorem [2].

Claim 1 $\lambda_n = \frac{1}{n}\mathbb{E}\{\max_{i,j}[\bigotimes_{k=1}^{n} M(k)]_{i,j}\}$ is a decreasing sequence bounded below by zero and therefore $\lim_{n\to\infty} \lambda_n = \lambda$ exists.

Proof Consider $(n+m)\lambda_{n+m}$ the expected weight of the maximally weighted path of length $n+m$. Now suppose that to further maximize this weight we are able to jump to any vertex in the graph between the nth and $(n+1)$th stages - since we are able to do nothing if we so choose the weight of this new path must be at least equal to the previous maximum so that

$$\begin{aligned}(n+m)\lambda_{n+m} = \mathbb{E}\{\max_{i,j}[\bigotimes_{k=1}^{n+m} M(k)]_{i,j}\} &\leq \mathbb{E}\{\max_{i,j,k,l}[\bigotimes_{k=1}^{n} M(k)]_{i,k} + [\bigotimes_{k=n+1}^{n+m} M(k)]_{l,j}\} \\ &= n\lambda_n + m\lambda_m\end{aligned} \tag{4.12}$$

and the claim follows from Fekete's Subadditive Lemma as in 2.1.3

Claim 2 $[\bigotimes_{k=1}^{n} M(k)]_{i,j} \geq \max_{i',j'}[\bigotimes_{k=1}^{n} M(k)]_{i',j'} - C$ for some C with bounded expectation.

proof Consider the path σ that attains the maximum weight in the right hand side of the inequality. Provided the graph associated with M is irreducible and aperiodic there exists a path ς from i to j that coincides with σ from the Nth step to the $n-N$th step. C is just the difference in weight between the two paths whose expectation can easily be bounded by $2N\lambda_1$.

Corollary Our theorem follows by considering a path of length $n \times m$ from i to j. Suppose that the maximally weighted path is σ then

$$\lim_{n,m\to\infty} \frac{1}{nm}[\bigotimes_{k=1}^{nm} M(k)]_{i,j} = \lim_{n,m\to\infty} \frac{1}{m}(\frac{1}{n}[\bigotimes_{k=1}^{n} M(k)]_{i,\sigma(n)} + \tag{4.13}$$

$$\frac{1}{n}[\bigotimes_{k=n+1}^{2n} M(k)]_{\sigma(n+1),\sigma(2n)} + ... + \frac{1}{n}[\bigotimes_{k=n(m-1)+1}^{nm} M(k)]_{\sigma(n(m-1)+1),j})$$

We can now use claim 2 to replace

$$\frac{1}{n}[\bigotimes_{k=tn+1}^{n(t+1)} M(k)]_{\sigma(tn+1),\sigma(n(t+1)} \tag{4.14}$$

with

$$\frac{1}{n}\max_{i',j'}[\bigotimes_{k=tn+1}^{n(t+1)} M(k)]_{i',j'} + \frac{C}{n} \tag{4.15}$$

Taking the limit in m the first part of each term gives us λ_n by the law of large numbers. Then taking the limit in n gives us λ and takes the C terms to zero. If we can not factorize our path length in this way simply include a remainder which when divided by n goes to zero.

Remark The proof for the special case of deterministic sequences is also simpler from this maximally weighted path perspective. Consider

$$\lambda = \lim_{n\to\infty} \frac{1}{n} A^{\otimes n} \tag{4.16}$$

where A is any irreducible aperiodic max-plus matrix. Now take the cycle $C = c(1), ..., c(m)$ with maximum average weight and consider the component

$A_{i,j}^{\otimes n}$ which is the weight of the maximally weighted path of length n from i to j. To construct a path attaining this weight (for large n) simply move from i to $c(1)$ in fewer than N (the number of vertices) steps, complete the cycle as many times as possible and finally return to j in less than m+N steps. The average weight of this and any other maximally weighted path will converge to C's average weight.

The dynamics of our max-plus system are therefore dominated by λ. The remainder of this chapter is devoted to calculating and bounding the exponent for some examples and developing a theory to link it with the classical eigenvalues of the adjacency matrix of M's associated graph.

4.2.1 Example

Since the max-plus exponent acts in an arithmetic way as opposed to a classical Lyapunov exponent's geometric action it is very easy to approximate directly in a numerically stable way. Returning to our asynchronous production line example we choose all the waiting times to be i.i.d. mean-1 exponentials and simulate the max-plus system as outlined in 4.1.3. The Max-plus exponent λ corresponds to the average time between successive completions of finished products in the production line, it is the reciprocal of the throughput.

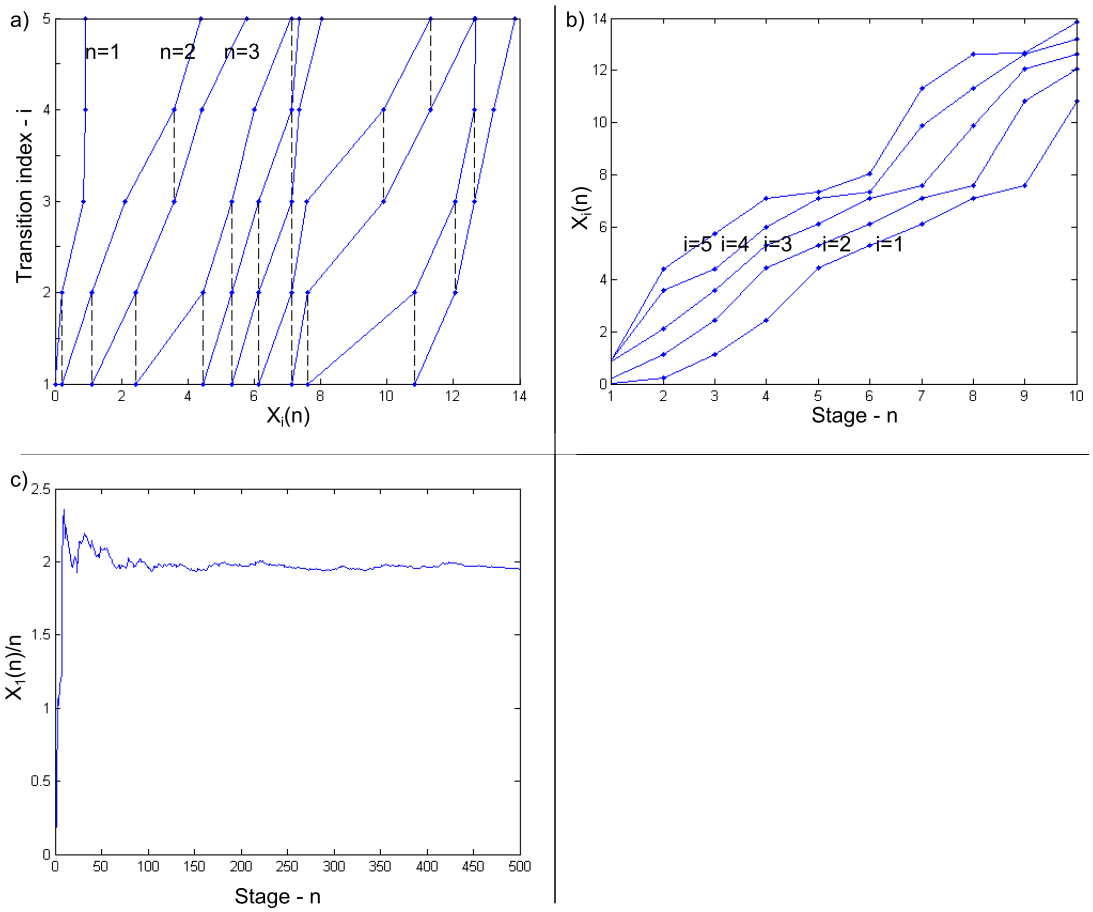


Figure 4.4: Simulation of asynchronous production line for $N = 4$. a) Progression of individual items through system, dashed black lines indicate jamming, b) Evolution of $X_i(n)$ variable, c) Convergence of $X_1(n)/n$ to max-plus Lyapunov exponent λ.

4.2.2 Relation with principal adjacency eigenvalue

The max-plus exponent λ depends on both the waiting-time distributions and the topology of the matrices associated graphs. By restricting our attention to matrices whose waiting times are all i.i.d. mean-1 exponentials we can better explore the relationship between λ and the graph's topology.

Definition we say that an $\mathbb{R}_{\max}^{N\times N}$ valued random variable M is *componentwise exponential* iff

$$M_{i,j} = \bigoplus_{k\in K(i,j)} t_k \tag{4.17}$$

Where $(t_k)_{k=1}^{L}$ is a sequence of i.i.d. mean-1 exponentials and $K(i,j) \subset \{1,2,...,L\}$. We will also require the second condition that $(t_{K(i,j)})_{j=1}^{N}$ are independent for each i.

Remark The graph associated with a componentwise exponential max-plus matrix is therefore a directed weighted graph on N vertices with possibly multiple edges between any two nodes. The weight of each edge is given by a mean-1 exponential random variable and the different edge's weights are *either* independent *or* identical. The second condition guarantees that all edges out of a particular vertex have independent weights. We can actually relax this condition further to allow the weights to be *associated* (as defined in the upper bound proof) provided that for each vertex the edge weights of its out edges are all independent.

Definition The *Adjacency matrix* A associated with a componentwise exponential matrix M is defined by

$$A(i,j) = |K(i,j)| \tag{4.18}$$

That is the number of independent time delays that contribute to $M(i,j)$. This needn't just be zero or one, it can be any integer, in terms of the graph a value greater than one corresponds to several parallel edges between two vertices.

Theorem Suppose that $[M(n)]_{n=1}^{\infty}$ is a sequence of i.i.d. componentwise exponential max-plus matrices. The max-plus exponent

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \bigotimes_{k=1}^{n} M(k) \tag{4.19}$$

satisfies

$$\log \Lambda \leq \lambda < \alpha^* \tag{4.20}$$

where α^* is the unique solution to

$$\alpha = \log \Lambda + \log(1 + e\alpha) \tag{4.21}$$

with $e = 2.817...$ and Λ the principal (classical) eigenvalue of $\mathcal{A}(M)$, the adjacency matrix of M's associated graph defied by $\mathcal{A}(M)_{i,j} = k$ iff there are k edges from j to i in M's graph.

Note The adjacency eigenvalue Λ tells us the rate at which the number of paths of length n through the graph grows with n. In particular if $\chi_i(n)$ is the number of paths of length n that end at vertex i then (in standard algebra!)

$$\chi(n+1) = \mathcal{A}(M)\chi(n) = \mathcal{A}(M)^{n+1}\chi(0) \tag{4.22}$$

Where $\chi(0) = (1, 1, ..., 1)$ and provided the adjacency matrix is irreducible and aperiodic there exists a unique maximal eigenvalue Λ (which we call the principal eigenvalue or Perron root) corresponding to a eigenvector u with non-zero weight on all components so that

$$\lim_{n\to\infty} \frac{1}{n} \log \|\chi(n)\| = \log \Lambda \tag{4.23}$$

Lower Bound Recall that

$$\lambda_n = \frac{1}{n}\mathbb{E}\{\max_{i,j}[\bigotimes_{k=1}^{n} M(k)]_{i,j}\} \tag{4.24}$$

is the expectation of the step averaged weight of the maximally weighted path of length n through the graphs associated with $[M(n)]_{n=1}^{\infty}$ and that $\lambda = \lim_{\to\infty} \lambda_n$. The max-plus exponent is therefore the step averaged weight of the maximally weighted path and we can bound it below with the step averaged weight of any other path we like. In particular we will devise a method for constructing a highly weighted path and use the expectation of its step averaged weight as a lower bound.

Our strategy constructs a path σ by choosing to move from one vertex to the next at each stage by considering only the edge weights at that stage. Suppose that at stage n our path is at vertex i so $\sigma(n) = i$. The edge weight to be accumulated on the next step will be taken from the matrix $M(n)$ and since we must move from vertex i it will be one of $M(n)_{i,j}$ for $j = 1, 2, ..., N$.

We will always greedily choose to move along the maximally weighted edge accumulating the maximum weight so that $\sigma(n+1) = k$ where $M(n)_{i,k} \geq M(n)_{i,j}$ for all $j = 1, 2, ..., N$.

Since the edges out of each vertex are all i.i.d. the probability that any particular edge has the maximum weight is the uniform fraction $1/d_i$ where d_i is the out degree of the vertex i. The sequence of vertices in the path σ is therefore a Markov chain with

$$\mathbb{P}[\sigma(n+1) = j | \sigma(n) = i] = \frac{A_{i,j}}{d_i} \tag{4.25}$$

and stationary distribution π.

Now consider the weight of the first n edges in the path σ

$$W|_n(\sigma) = \max_{k \in K[\sigma(1)]} t_k(1) + \max_{k \in K[\sigma(2)]} t_k(2) + ... + \max_{k \in K[\sigma(n)]} t_k(n) \tag{4.26}$$

where $K[i] = \cup_{j \in \{1,2,...,N\}} K(j,i)$ are the indices for all waiting-times associated with edges leaving vertex i. By the law or large numbers we have

$$\lim_{n \to \infty} \frac{1}{n} W|_n(\sigma) = \sum_{i=1}^{N} \pi_i \underbrace{\mathbb{E}[\max_{k \in K(i)} t_k]} \tag{4.27}$$

where the underbraced term is the expectation of the maximum of d_i i.i.d. mean-1 exponentials.

Claim The expectation of the maximum of d i.i.d. mean-1 exponentials $t_1, ..., t_d$ equals $\sum_{k=1}^{d} \frac{1}{k}$.

Proof Suppose that each exponential variable is the waiting-time from $t = 0$ for some event to occur. The first event is the minimum of d exponentials, which is a mean-$\frac{1}{d}$ exponential. Once this event has occurred we can apply the memory loss property of the remaining exponential distributions so that the time between the first and second events is the minimum of $d-1$ exponentials, which is a mean-$\frac{1}{d-1}$ exponential. Continuing in this way the time of dth event (the maximum of the d mean-1 exponentials) is the sum of d independent exponentials with

means $\frac{1}{d}, \frac{1}{d-1}, ..., \frac{1}{2}, 1$ and by independence the mean of the sum is the sum of these means.

The step averaged weight of the path is therefore given by

$$\sum_{i=1}^{N} \pi_i \sum_{k=1}^{d_i} \frac{1}{k} \geq \sum_{i=1}^{N} \pi_i \log d_i \tag{4.28}$$

where we will use the log approximation as our lower bound.

We now consider a cocycle on the Markov chain σ defined by

$$C|_n(\sigma) = [\prod_{k=0}^{n} d_{\sigma(k)}]^{-1} \tag{4.29}$$

and by the multiplicative ergodic theorem, with probability one, we have

$$\lim_{n\to\infty} \frac{1}{n} \log C|_n(\sigma) = -\sum_{i=1}^{N} \pi_i \log d_i \tag{4.30}$$

The nth stage entropy of the cocyle is given by

$$H_n = \frac{-1}{n} \sum_{\{\sigma|_n : \sigma(0)=1\}} C|_n \log C|_n \tag{4.31}$$

where the sum is taken over all paths of length n that begin at vertex 1. Note that $C|_n(\sigma)$ is the probability of following exactly the path σ from $\sigma(0)$ to $\sigma(n)$ taking the appropriate edge $[\sigma(k+1), \sigma(k)]$ whenever their are multiple edges, we therefore say that our cocycle is stochastic and

$$\lim_{n\to\infty} H_n = -\sum_{\{\sigma|_n : \sigma(0)=1\}} C|_n(-) \sum_{i=1}^{N} \pi_i \log d_i = \sum_{i=1}^{N} \pi_i \log d_i \tag{4.32}$$

The entropy of a probability measure on a set of k objects is maximized when the measure is uniform. Since $C|_n$ can be thought of as a probability measure on $\{\sigma|_n : \sigma(0) = 1\}$ and in the limit $n \to \infty$ C_n becomes, essentially, uniform we have

$$\lim_{n\to\infty} H_n - \frac{1}{n}\widehat{H}_n = 0 \tag{4.33}$$

where $\widehat{H}(n)$ is the entropy of the uniform probability distribution on $\{\sigma|_n : \sigma(0) = 1\}$ which is given by

$$\widehat{H}(n) = -\sum_{\{\sigma|_n : \sigma(0)=1\}} \frac{1}{D_n} \log D_n \tag{4.34}$$

where $D_n = \|\{\sigma|_n : \sigma(0) = 1\}\|$ is the number of paths of length n that start at vertex 1. Finally we have

$$\sum_{i=1}^{N} \pi_i \log d_i = \lim_{n\to\infty} \frac{1}{n} \log D_n = \log \Lambda \tag{4.35}$$

which completes the proof of our lower bound

$$\log \Lambda < \lambda \tag{4.36}$$

Remark The use of the approximation

$$\sum_{k=1}^{N} \frac{1}{k} \approx \log N \tag{4.37}$$

is the reason that our bound is not sharp. This is in the sense that we can construct a sequence of i.i.d. matrices whose maximally weighted path is the same path chosen by our greedy strategy. If we set

$$M(n)_{i,j} = t_i(n) \tag{4.38}$$

for $i, j = 1, 2, ..., N$ then the choice of edge weights at each vertex are the same for all the different vertices at each stage and the optimal strategy for choosing a maximally weighted path is simply to choose the best available edge at each stage, as in our greedy strategy. Therefore in this system

$$\lambda = \sum_{k=1}^{N} \frac{1}{k} \tag{4.39}$$

and we can express this in terms of $\Lambda = N$ (the number of vertices) in the limit by

$$\lim_{N\to\infty} \lambda = \log \Lambda + \gamma + \epsilon_N \tag{4.40}$$

where γ is the EulerMascheroni constant and $\epsilon_N \to 0$ like $1/N$.

Upper bound The expected weight of the maximally weighted path of length n can, of course, be expressed in terms of path weights

$$\lambda_n = \frac{1}{n}\mathbb{E} \max_{i,j} [\bigotimes_{k=1}^{n} M(k)]_{i,j} = \frac{1}{n}\mathbb{E} \max_{\sigma|_n} \underbrace{W|_n(\sigma, [t_i(k)]_{i=1,k=1}^{i=N,k=n})}_{} \tag{4.41}$$

where the underbraced term is equal to the weight of the length-n path σ with edge weights determined by the i.i.d. mean-1 exponentials $[t_i(k)]_{i=1,k=1}^{i=N,k=n}$. It is important that in evaluating this expression the same waiting times are used for each different path. The maximum is then taken over all paths of length n.

The weight of each path of length-n is a sum of n i.i.d. mean-1 exponentials so that all path weights are identically distributed. However different paths can share edges which they traverse at the same step, accumulating the same waiting time. Therefore the path weights are not independent. Since their dependence arises from edge weight sharing it will only tend to make them more correlated. We say that the edge weights are *associated* random variables.

Definition A sequence of random variables $(x_i)_{i=1}^N$ are said to be *associated* if for all $f, g : \mathbb{R}^N \mapsto \mathbb{R}$ non-decreasing in each component we have

$$Cov[f(x_1, x_2, ..., x_N), g(x_1, x_2, ..., x_N)] \geq 0 \tag{4.42}$$

To see that the path weights are associated note that each $W|_n(\sigma)$ is a non-decreasing function of the waiting-times. Thus if f and g are non-decreasing functions of the path weights they are non-decreasing functions of the waiting-times which are all independent so the inequality holds.

Since this association means the random variables are positively correlated it reduces their standard deviation and the expectation of their maximum. We can therefore bound λ from above by taking the maximum in (4.41) while ignoring the dependence and treating each path weight as an i.i.d. sum of i.i.d. mean-1 exponentials.

Claim Suppose that $(x_i)_{i=1}^N$ and $(y_i)_{i=1}^N$ are identically distributed random variables and that $(x_i)_{i=1}^N$ are associated but $(y_i)_{i=1}^N$ are independent.

$$\mathbb{E} \max_{i=1}^{N} x_i \leq \mathbb{E} \max_{i=1}^{N} y_i \tag{4.43}$$

Proof For any $i \in \{1, 2, ..., N\}$, any subset $J \in \{1, 2, .., N\}$ and any positive real

number t define the non-decreasing functions f and g by

$$f(x_1, x_2..., x_N) = \begin{cases} -1 & \text{if } x_i < t \\ 0 & \text{otherwise} \end{cases} \tag{4.44}$$

$$g(x_1, x_2..., x_N) = \begin{cases} -1 & \text{if } \max_{j\in J} x_j < t \\ 0 & \text{otherwise} \end{cases}$$

Now to calculate the covariance of these two functions we need to consider four events

- $x_i < t$ and $\max_{j\in J} x_j < t$ occurs with probability $P(f,g) = \mathbb{P}[x_i < t, x_j < t, j \in J]$
- $x_i < t$ but $\max_{j\in J} x_j \geq t$ occurs with probability $P(f) - P(f,g)$ where $P(f) = \mathbb{P}[x_i < t]$
- $\max_{j\in J} x_j < t$ but $x_i \geq t$ occurs with probability $P(g) - P(f,g)$ where $P(g) = \mathbb{P}[x_j < t, j \in J]$
- $x_i \geq t$ and $\max_{j\in J} x_j \geq t$ occurs with probability $1-P(f)-P(g)+P(f,g)$

The covariance of f, g is then

$$Cov[f,g] = P(f,g)[P(f)P(g)-P(f)-P(g)+1]+[P(f)-P(f,g)][P(f)-1]P(g) \tag{4.45}$$
$$+[P(g)-P(f,g)][P(g)-1]P(f) + [1-P(f)-P(g)+P(f,g)]P(f)P(g) \geq 0$$

which simplifies to $P(f,g) \geq P(f)P(g)$ so that

$$\mathbb{P}[x_i < T | x_j < T, j \in J] \geq \mathbb{P}[x_i < T] \tag{4.46}$$

Now consider

$$\mathbb{P}[\max_{i=1}^{N} x_i < t] = \prod_{i=1}^{N} \mathbb{P}[x_i < t | x_k < t; k = 1, 2, ..., i-1] \tag{4.47}$$

and if we set $J = \{1, 2, ..., i-1\}$ then our previous result tells us that

$$\mathbb{P}[\max_{i=1}^{N} x_i < t] \geq \prod_{i=1}^{N} \mathbb{P}[x_i < t] \tag{4.48}$$

which is exactly equal to the probability that the maximum of the y_i is less than t. Finally

$$\mathbb{E}\max_{i=1}^{N} x_i = \int_0^\infty t\rho_x(t)dt = \int_0^\infty \int_z^\infty \rho_x(t)dtdz \tag{4.49}$$

$$= \int_0^\infty 1 - \mathbb{P}[\max_{i=1}^{N} x_i < z]dz \leq \int_0^\infty 1 - \mathbb{P}[\max_{i=1}^{N} y_i < z]dz = \mathbb{E}\max_{i=1}^{N} y_i$$

so that the expectation of the maximum of the associated variables is less than or equal to the expectation of the maximum of the independent variables as claimed.

Therefore we have the upper bound

$$\lambda \leq \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\max_{i=1}^{D_n} X_i(n) \tag{4.50}$$

where $[X_i(n)]_{i=1}^{D_n}$ is a sequence of i.i.d. sums of n i.i.d. mean-1 exponentials and D_n is the number of paths of length n through our graph. Unfortunately we can't approximate the $X_i(n)$ as Gaussians because the convergence in the central limit theorem is not sufficiently uniform. Instead we will show directly that as $n \to \infty$ the ratio of the median to the mean of the maximum tends to 1 so we can calculate the expectation by calculating the median, which is naturally easier than the mean for the maximum of several random variables.

Given a sequence of probabilities $(p_n)_{n=1}^\infty$ we define the sequence of generalized medians $[\mu_n(p_n)]_{n=1}^\infty$ for our sequence of random variables $[\max_{i=1}^{D_n} X_i(n)]_{n=1}^\infty$ by

$$\mathbb{P}[\max_{i=1}^{D_n} X_i(n) < \mu_n(p_n)] = \mathbb{P}[X_1(n) < \mu_n(p_n)]^{D_n} = p_n \tag{4.51}$$

And using the identity

$$e^x = \lim_{m\to\infty}(1 + \frac{x}{m})^m \tag{4.52}$$

with $x = \log p_n$ and $m = D_n$ we have, approximately but verifiably in the limit $n \to \infty$ for all sequences $(p_n)_{n=1}^\infty$ we consider

$$\mathbb{P}[X_1(n) < \mu_n(p_n)] = 1 + \frac{\log p_n}{D_n} \tag{4.53}$$

$$\mathbb{P}[X_1(n) > \mu_n(p_n)] = -\frac{\log p_n}{D_n}$$

And since we know the distribution of the sum of n i.i.d. mean-1 exponentials we can write down an integral for the LHS which we can then integrate by parts

$$= \int_{\mu_n(p_n)}^{\infty} \frac{x^{n-1}e^{-x}}{(n-1)!}dx = e^{-\mu_n(p_n)} \underbrace{\sum_{k=0}^{n-1} \frac{\mu_n(p_n)^k}{k!}} \tag{4.54}$$

where the underbraced term is a truncation of e^x's power series evaluated at $x = \mu_n(p_n)$. We can bound this sum with the inequality

$$(1 + \frac{x}{n})^n \leq \sum_{k=0}^{n} \frac{x^k}{k!} \leq (1 + \frac{ex}{n})^n \tag{4.55}$$

which we prove by looking at the coefficient of x^k

$$\frac{1}{k!}\frac{n!}{(n-k)!n^k} \leq \frac{1}{k!} \leq \frac{1}{k!}\frac{n!e^k}{(n-k)!n^k} \tag{4.56}$$

Therefore in the limit $n \to \infty$ we have the inequality

$$e^{-\mu_n(p_n)}(1 + \frac{\mu_n(p_n)}{n-1})^{n-1} \leq \frac{-\log p_n}{D_n} \leq e^{-\mu_n(p_n)}(1 + \frac{e\mu_n(p_n)}{n-1})^{n-1} \tag{4.57}$$

Claim Any sequence of medians $\mu_n(p_n)$ satisfying

$$\lim_{n\to\infty} \frac{\log\log p_n}{n} = 0 \tag{4.58}$$

including the sequence of proper medians $[\mu_n = \mu_n(\frac{1}{2})]_{n=1}^{\infty}$ satisfy, in the limit $n \to \infty$

$$\log\Lambda \leq \frac{\mu_n}{n} \leq \log\Lambda + \log(1 + \frac{e\mu_n}{n}) \leq C\log\Lambda + D \tag{4.59}$$

for some bounded C D that do not depend on Λ.

Proof substituting $p_n = \frac{1}{2}$ into inequality (4.57) then taking logs and dividing by n gives the first sandwich inequality. That everything is less than or equal to $C\log\Lambda + D$ follows from the fact that

$$F(\alpha) = \log\Lambda + \log(1 + e\alpha) \tag{4.60}$$

is a contraction mapping whose unique fixed point α^* bounds $\lim_{n\to\infty} \frac{\mu_n}{n}$ from above. Now

$$\frac{\delta\alpha^*}{\delta\log\Lambda} = \frac{1 + e\alpha^*}{1 + e(\alpha^* - 1)} \tag{4.61}$$

So that for any fixed finite Λ' we have

$$\alpha^*(\Lambda) \leq \alpha^*(\Lambda') + \frac{\delta\alpha^*}{\delta \log \Lambda}|_{\alpha^*}(\Lambda') \log \Lambda \tag{4.62}$$

Claim For large n almost all the probability mass is close to the median in the sense that

$$\lim_{n\to\infty} \frac{\mu_n(e^{\frac{-1}{n}}) - \mu_n(e^{-n})}{n} = 0 \tag{4.63}$$

Proof Substituting $p_n = e^{\frac{-1}{n}}$ and e^{-n} into equation (4.54) then taking logs we have

$$\frac{\mu_n(e^{\frac{-1}{n}})}{n} = \log \Lambda + \frac{\log n}{n} + \log(\sum_{k=0}^{n-1} \frac{\mu_n(e^{\frac{-1}{n}})^k}{k!}) \tag{4.64}$$

and

$$\frac{\mu_n(e^{-n})}{n} = \log \Lambda - \frac{\log n}{n} + \log(\sum_{k=0}^{n-1} \frac{\mu_n(e^{-n})^k}{k!}) \tag{4.65}$$

respectively, So that

$$\frac{\mu_n(e^{\frac{-1}{n}}) - \mu_n(e^{-n})}{n} \leq \frac{2\log n}{n} \tag{4.66}$$
$$+\frac{\mu_n(e^{\frac{-1}{n}}) - \mu_n(e^{-n})}{n} \underbrace{\max_{\mu_n(e^{-n}) \leq x \leq \mu_n(e^{\frac{-1}{n}})} \frac{d}{dx} \log(\sum_{k=0}^{n-1} \frac{x^k}{k!})}$$

using the result of the previous claim we bound the underbraced term by

$$\max_{\log\Lambda \leq \alpha \leq C\log\Lambda + D} 1 - \frac{\frac{(\alpha n)^{n-1}}{(n-1)!}}{\sum_{k=0}^{n-1} \frac{(\alpha n)^k}{k!}} \tag{4.67}$$

which gives

$$\frac{\mu_n(e^{\frac{-1}{n}}) - \mu_n(e^{-n})}{n} \leq \frac{2\log n}{n} \underbrace{\max_{\log\Lambda \leq \alpha \leq C\log\Lambda + D} \frac{(n-1)!}{(\alpha n)^{n-1}} \sum_{k=0}^{n-1} \frac{(\alpha n)^k}{k!}} \tag{4.68}$$

where the underbraced term can be bounded above by

$$\max_{\log\Lambda \leq \alpha \leq C\log\Lambda + D} \sum_{k=0}^{n-1} \frac{1}{\alpha^{n-k-1}} \leq \max_{\log\Lambda \leq \alpha \leq C\log\Lambda + D} (1 - \frac{1}{\alpha})^{-1} \tag{4.69}$$

So that

$$\frac{\mu_n(e^{\frac{-1}{n}}) - \mu_n(e^{-n})}{n} < \frac{2\log n}{n}(1 - \frac{1}{\log\Lambda})^{-1} \to 0 \tag{4.70}$$

as required.

We can now express the mean as an integral over a series of intervals bounded by generalized medians

$$\mathbb{E}\max_{i=1}^{D_n} X_i = \int_0^{\mu_n(e^{-n})} x\rho(x)dx + \underbrace{\int_{\mu_n(e^{-n})}^{\mu_n(e^{\frac{-1}{n}})} x\rho(x)dx}$$

$$+ \int_{\mu_n(e^{\frac{-1}{n}})}^{\mu_n(e^{\frac{-1}{m^n}})} x\rho(x)dx + \sum_{m=2}^{\infty} \int_{\mu_n(e^{\frac{-1}{m^n}})}^{\mu_n(e^{\frac{-1}{(m+1)^n}})} x\rho(x)dx \tag{4.71}$$

Where the underbraced term is equal to $\mu_n(\frac{1}{2}) + o(n)$. All that remains is to show that the sum of the remaining terms grows slower than n.

The first integral is easy

$$\int_0^{\mu_n(e^{-n})} x\rho(x)dx \leq e^{-n}\mu_n(e^{-n}) \leq e^{-n}n(C\log\Lambda + D) \tag{4.72}$$

We treat the remaining integrals in the same way, bounding them above by the probability associated with their interval multiplied by an upper bound on the value of their upper boundary. Using the upper bound in inequality (4.57) we obtain

$$\frac{\mu_n(e^{\frac{-1}{m^n}})}{n} < \log\Lambda - \frac{1}{n}\log\log e^{\frac{1}{m^n}} + \log(1 + e\frac{\mu_n(e^{\frac{-1}{m^n}})}{n}) \tag{4.73}$$

So that

$$\frac{\mu_n(e^{\frac{-1}{m^n}})}{n} < \log m\Lambda + \log(1 + e\frac{\mu_n(e^{\frac{-1}{m^n}})}{n}) \tag{4.74}$$

and as in (4.62) we have

$$\frac{\mu_n(e^{\frac{-1}{m^n}})}{n} < C\log m\Lambda + D \tag{4.75}$$

So the sum of the remaining integrals is bounded by

$$n(e^{\frac{-1}{2^n}} - e^{\frac{-1}{n}})(C\log 2\Lambda + D) + n\sum_{m=3}^{\infty}(e^{\frac{-1}{m^n}} - e^{\frac{-1}{(m-1)^n}})[C\log mK(n) + D] \tag{4.76}$$

And in the limit $n \to \infty$ this expression is bounded above by

$$(1 - e^{-n})n(C\log 2\Lambda + D) + n\sum_{m=3}^{\infty}\frac{C\log m}{(m-1)^n} \tag{4.77}$$

$$\leq Cn\sum_{m=3}^{\infty}\frac{1}{(m-1)^{n-1}} \to 0$$

Therefore the mean really does look like the median and we can use the bound obtained for the median on the mean.

Corollary Suppose that $[M_N(n)]_{n=1}^{\infty}$ is a sequence of i.i.d. componentwise exponential max-plus matrices parameterized by $N \in \mathbb{N}$.

- If $\max_N \Lambda_N = \overline{\Lambda}$ exists then $\max_N \lambda_N = \overline{\Lambda}$ exists.
- If on the other hand $\Lambda_N \to \infty$ then $\lim_{N\to\infty} \frac{\lambda_N}{\log \Lambda_N} = 1$
- Moreover if, as in all our examples, $M_N(n)$ can be realized as submatrix of $M_{N+K}(n)$ for all $N, K \geq 0$ then either $\lim_{N\to\infty} \Lambda_N$ exists and so does $\lim_{N\to\infty} \lambda_N$ or $\lim_{N\to\infty} \frac{\lambda_N}{\log \Lambda_N} = 1$

Proof The final item comes from the fact that if M_N is a submatrix of M_{N+K} then any path through its associated graph can also be found in the graph of M_{N+K} so that $\Lambda_N \leq \Lambda_{N+K}$ and $\lambda_N \leq \lambda_{N+K}$. Everything else follows directly from the main theorem.

4.3 Examples

Our result relating the max-plus exponent to the adjacency matrix eigenvalue is restrictive in the sense that it only applies to componentwise exponential matrices. The max-plus matrix product systems associated with stochastic event graphs are not typically in this form. For example in 4.1.3 the graph associated with $A^*(n)B(n-1)$ has edges with zero weight as well as edges whose weight are conventional sums of more than one i.i.d. mean-1 exponential.

However we shall see that this and other matrices associated with a generalization of our original production line model can be modified in such a way that preserves the max-plus exponent and provides us with a componentwise exponential form. Thus we can apply our *edge weight redistribution* procedure to obtain a new system which has the same asymptotic properties as the original stochastic event graph but is amenable to our new theory.

4.3.1 Asynchronous production trees

The production line model outlined in 4.1.1 requires the N different process to be carried out in the total order $1 > 2 > ... > N$ to produce a finished product. We can generalize our production model to allow for a class of partial orders on these processes.

Definition A *tree order* $\preceq, \succeq$ is a partial order on $\{1, 2, ..., N\}$ with a unique maximal element such that $\{k \preceq i\} \cap \{k \preceq j\} = \emptyset$ whenever $\{i, j\}$ is an anti-chain. We associate the graph $G(\preceq, \succeq)$ with vertex set $\{1, 2, ..., N\}$ and an edge (i, j) whenever j is a maximal element of $\{k \preceq i\}$. The graph associated with a tree metric is therefore a rooted tree and for each rooted tree on $\{1, 2, ..., N\}$ there exists a unique partial order associated with it as such.

The requirement that incomparable processes do not share predecessors is necessary to obtain a stochastic event graph description as otherwise some places will have more than one succeeding transition. It is also a fairly reasonable assumption for a generic asynchronous production line, if we interpret each process as taking the products of its direct predecessors and amalgamating them in some way then there is no reason why parallel processes should be dependent in any way on each others predecessors.

We construct the stochastic event graph associated with a tree order $\preceq, \succeq$ as follows. For each $i \in \{1, 2, ..., N\}$ we include a pair of places p_i and q_i, and a transition T_i. As before a token in p_i represents as an item being processed at site i with i.i.d. mean-1 exponential waiting times $[t_i(n)]_{n=1}^{\infty}$. A token in q_i represents site i being unoccupied with zero waiting times. Transition T_i represents the products of process i's predecessors being completed and moved into site i for processing.

We include an edge from q_i to T_i and from T_i to p_i, in addition we include an edge from p_i to T_j where j is i's unique direct successor in $G(\preceq, \succeq)$ and an edge from T_j to q_i.

Finally we include a terminal transition T_{N+1} which represents the completion of a finished product, we include a edge from T_{N+1} to q_k where k is the root of $G(\preceq, \succeq)$ and an edge from p_k to T_{N+1}. See example 4.3.4.

4.3.2 Edge weight redistribution

The graphs associated with $A(n)$ and $B(n-1)$ for the stochastic event graph of an asynchronous production line with partial order $G(\preceq, \succeq)$ are as follows.

- $A(n)$ is a graph on $\{1, 2, ..., N+1\}$ with an edge (i, j) of weight $t_j(n)$ whenever (i, j) is an edge in $G(\preceq, \succeq)$, in addition there is an edge $(N+1, k)$ of weight $t_k(n)$ where k is the maximal element.

- $B(n-1)$ has the same vertex set but with the opposite edge set, so that (i, j) is an edge of weight zero in $B(n-1)$ whenever (j, i) is an edge in $A(n)$.

- $A^*(n)B(n-1)$ is a graph on $\{1, 2, ..., N+1\}$ where for each i there are self loops (i, i) of weight $t_j(n)$ and zero weight edges (j, i) for each j that is directly succeeded by i and for each $k \succeq i$ there is an edge of weight $t_j(n) + W_{k,i}$ where the second term is the weight of the unique path from i to k.

Definition $\widehat{A}(n)$ is the componentwise exponential matrix whose graph on $\{1, 2, ..., N+1\}$ contains for each i edges (j, i) and (i, k) of weight $t_j(n)$ for each j a direct predecessor of i and any $k \succeq i$. The matrices components are given by

$$\widehat{A}(n)_{i,j} = \begin{cases} t_i & \text{if } j \text{ is a direct predecessor of } i \\ \bigoplus_{k \in P(i)} t_k & \text{if } j \succeq i \\ \epsilon & \text{otherwise} \end{cases} \tag{4.78}$$

where $P(i)$ is the set of i's direct predecessors.

Claim The max-plus exponent of the edge weight redistributed system

$$\widehat{\lambda} = \lim_{n \to \infty} \frac{1}{n} [\bigotimes_{k=1}^{n} \widehat{A}(k)]_{N+1,N+1} \tag{4.79}$$

is equal to that of the original stochastic event graph system

$$\lambda = \lim_{n \to \infty} \frac{1}{n} [\bigotimes_{k=1}^{n} A^*(k)B(k-1)]_{N+1,N+1} \tag{4.80}$$

Note that we will be calculating these exponents by looking just at the $(N+1, N+1)$th component of the product which should be thought of as the averaged

weight of the maximally weighted path of length n from the maximal vertex to itself.

Proof Again we can express the exponent in terms of path weights.

$$\widehat{\lambda}_n = \frac{1}{n}\mathbb{E}\max_{i,j}[\bigotimes_{k=1}^{n}\widehat{A}(k)]_{i,j} = \frac{1}{n}\mathbb{E}\max_{\widehat{\sigma}|_n}\underbrace{W|_n(\widehat{\sigma},[\widehat{t_i}(k)]_{i=1,k=1}^{i=N,k=n})}\tag{4.81}$$

and since the $\widehat{t_i}(n)$ and $t_j(m)$ are all i.i.d. we can obtain a statistically valid description of this system by setting $\widehat{t_i}(n) = t_i[n + d(i)]$ where $d(i)$ is the depth of i in the partial order, i.e. the length of the longest chain from the maximal element to i. We shall now show that there is a map between the sets of paths such that any sequence of edge weights in the original system can be accumulated by a different path in the redistributed system and that therefore the two exponents are the same.

Paths from vertex $N + 1$ to vertex $N + 1$ in the original system must step down the tree one vertex at a time and are then able to jump back up. In the redistributed system paths jump down the tree then step back up. We define a map R from the paths in the original system to paths in the redistributed system by reading σ and simultaneously writing $R(\sigma)$. Let i be the latest step to be read from σ and j the latest step written to $R(\sigma)$.

- Start with $R(\sigma)_1 = \sigma_1 = N + 1$
- Whenever σ steps down the tree so that $d(\sigma_i) < d(\sigma_{i+1}) < ... < d(\sigma_{i+k}) \geq d(\sigma_{i+k+1})$ we set $R(\sigma)_j = \sigma_i$ $R(\sigma)_{j+1} = \sigma_{i+k}$
- whenever σ jumps back up the tree (including when it stays at the same level) so that $d(\sigma_i) \geq d(\sigma_{i+1})$ we set $R(\sigma)_j = \sigma_i$, $R(\sigma)_{j+1} = \sigma_i^{(1)}$, ...,$R(\sigma_{j+k}) = \sigma_i^{(k)} = \sigma_{i+1}$ where $\sigma_i^{(t)}$ is σ_i's unique tth successor.

Therefore if σ is a length n path from $N + 1$ to $N + 1$ in the original graph $R(\sigma)$ is a length n path from $N + 1$ to $N + 1$ in the redistributed graph, and

$$\widehat{W}R(\sigma) = W(\sigma)\tag{4.82}$$

So that we have a measure preserving map between the $\widehat{t}_i(n)$ and $t_j(m)$ and a bijection between the paths in the maximum. Therefore when we take the expectation we can integrate over the $t_j(m)$ and have an identical integrand for each exponent. Thus the exponents are the same as claimed.

4.3.3 Asynchronous production line

We have already calculated $A(n), B(n-1)$ and $A^*(n)B(n-1)$ for the asynchronous production line in section 4.1.3. Following the definition in 4.3.2 we construct the redistributed componentwise exponential matrix $\widehat{A}$ for this system

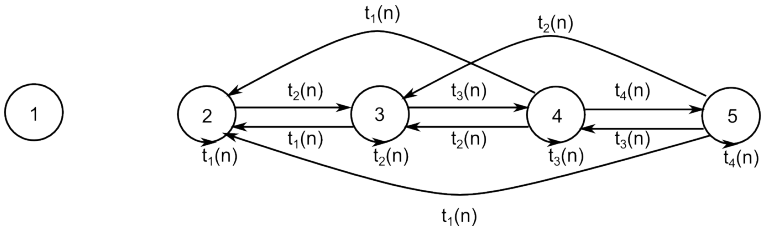


Figure 4.5: Graph associated with $\widehat{A}(n)$ for $N = 4$

Ignoring the minimal vertex with no connections the adjacency matrix for this graph is given by

$$\mathcal{A}_{i,j} = \begin{cases} 1 & \text{if } i \leq j+1 \\ 0 & \text{otherwise} \end{cases} \tag{4.83}$$

where we have identified each vertex with its depth for the index in the adjacency matrix. Now $\mathcal{A}$ is irreducible and aperiodic so Λ exists and is unique. Also there exists u with $u_i > 0$ for all i such that $\mathcal{A}u = \Lambda u$. Examining the first row of this equation gives $u_1 + u_2 = \Lambda u_1$ so that $u_2 = \Lambda - 1$ the nth row says

$$\sum_{j=1}^{n+1} u_j = \Lambda u_{n-1} + u_{n+1} = \Lambda u_n \tag{4.84}$$

Which gives us a second-order, linear, constant-coefficient recurrence relation on the u_n that we use to obtain

$$u_n = A_+ \mu_+^{n-1} + A_- \mu_-^{n-1} \tag{4.85}$$

where

$$\mu_{\pm} = \frac{\Lambda \pm \sqrt{\Lambda^2 - 4\Lambda}}{2} \tag{4.86}$$

and

$$A_{\pm} = \frac{1}{2} \pm \frac{\Lambda - 2}{2\sqrt{\Lambda^2 - 4\Lambda}} \tag{4.87}$$

We now require $u_N = u_{N-1}$ in order for the eigenvalue/vector pair to be valid. Suppose that $\Lambda > 4$ then both exponents are real and positive with $\mu_+ > \mu_-$ also $A_+ > |A_-|$ so that u_n is an increasing function of n and it is impossible to satisfy the final condition. Therefore $\Lambda \leq 4$ which proves that there is a finite limit in the exponent as the length of the production line grows $N \to \infty$. In fact $\Lambda \to 4$ very quickly and this can be seen in the convergence of the upper and lower bounds with N.

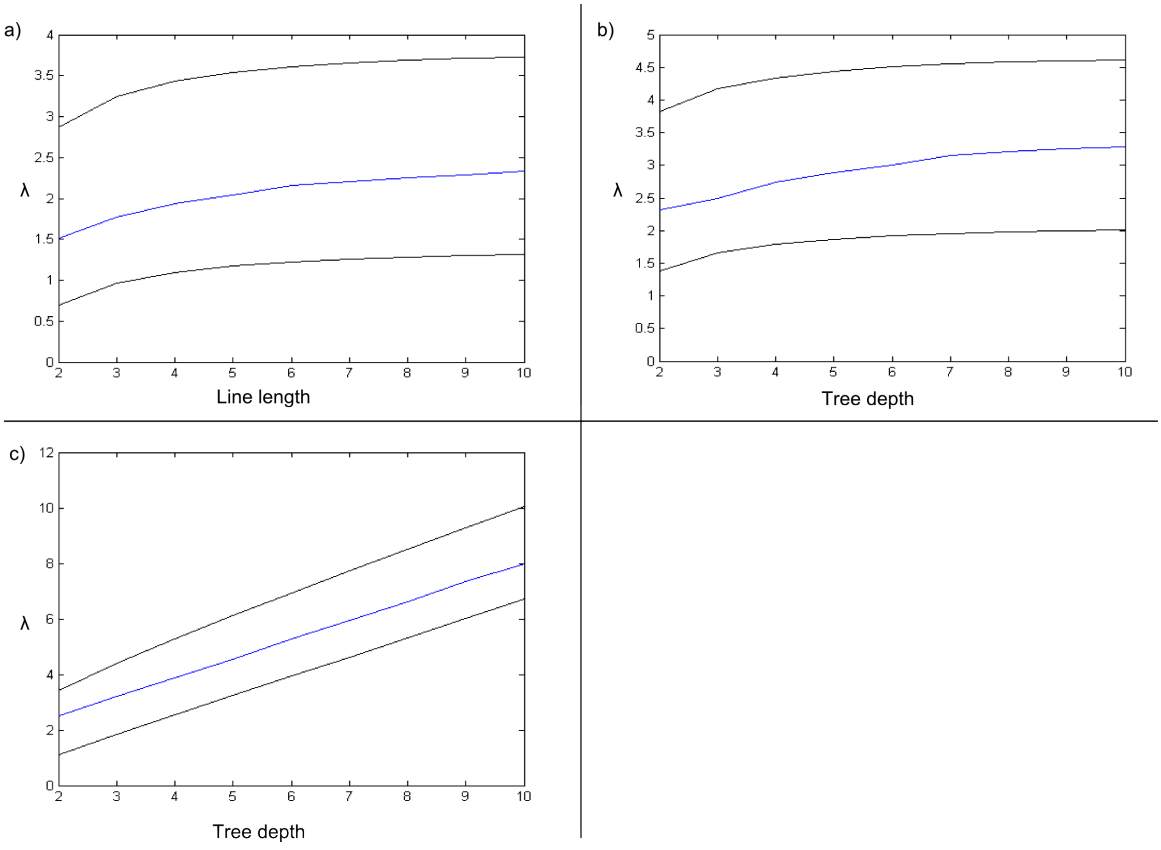


Figure 4.6: Max-plus exponent of production trees (blue) with bounds (black). a)Asynchronous production line, b)Asynchronous production binary tree, c)Partially synchronous production tree

4.3.4 Asynchronous production binary tree

We can construct an asynchronous production tree around a partial order derived from a binary tree.

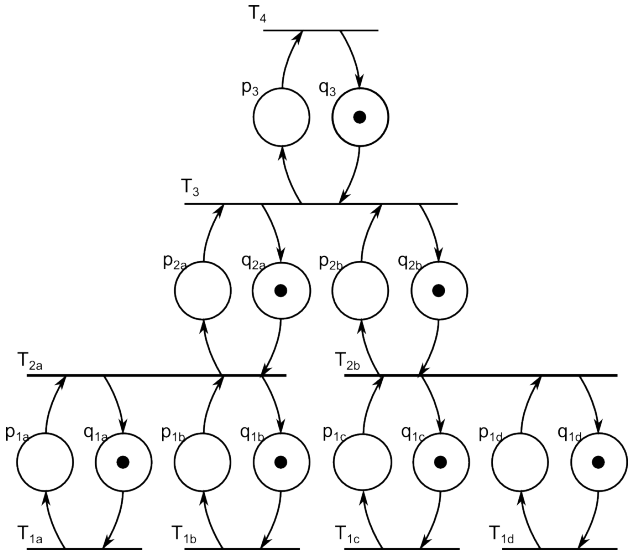


Figure 4.7: Petri-net and initial marking for asynchronous production binary tree for $N = 3$.

Calculation of the matrix product system and redistributed componentwise exponential matrix is as outlined previously.

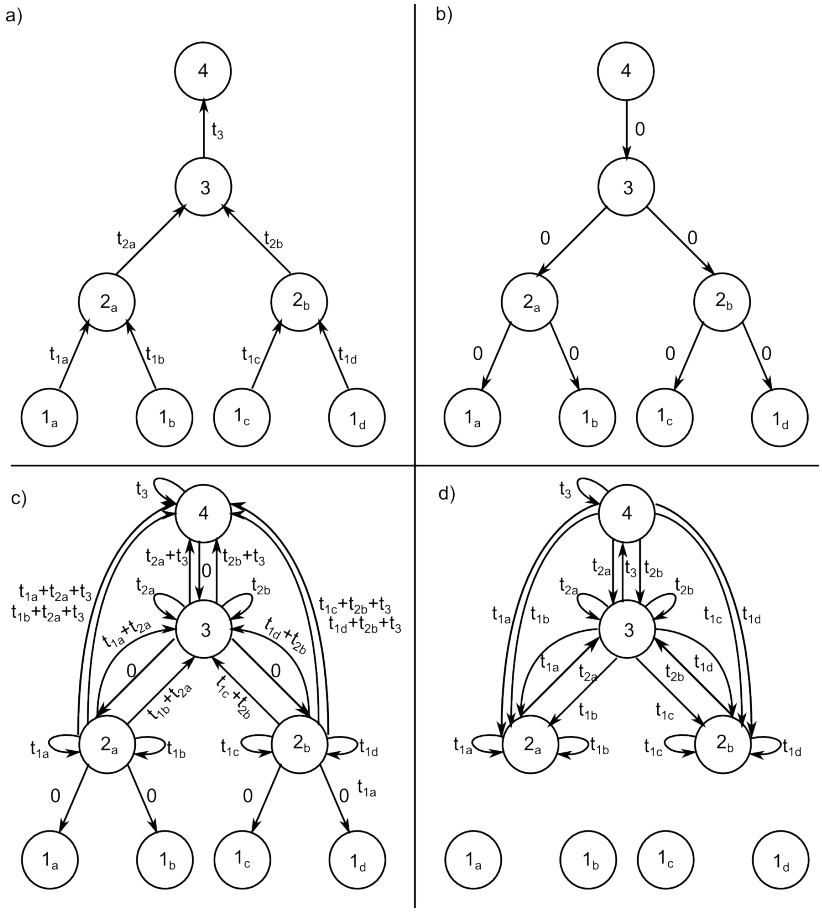


Figure 4.8: Graphs associated with A, B, A^*B and $\widehat{A}$ respectively for $N = 3$.

To compute the adjacency eigenvalue of the graph for $\widehat{A}$ we first quotient the vertex set so that we consider sets of depth i vertices. The adjacency matrix for this quotient is then given by

$$\mathcal{A}_{i,j} = \begin{cases} 1 & \text{if } i = 1,\ j = 1, 2 \\ 2 & \text{if } j \leq i + 1 \geq 3 \\ 0 & \text{otherwise} \end{cases} \tag{4.88}$$

which will have the same principal eigenvalue as the whole adjacency matrix. Note that this adjacency matrix is less than (componentwise) the adjacency matrix associated with the production line of length N multiplied by 2. Therefore the apparent asymptotes in the bound and the exponent definitely do exist.

4.3.5 Partially synchronous production binary tree

We can generalize our production line model further by requiring some anti-chains of processes to be completed synchronously. In terms of the production line model this means demanding that some processes that do not share any precursors must start at exactly the same time, so that one process may have to wait for the other to be ready in order for it to begin. Formally we take a partition of $\{1, 2, ..., N\}$ into disjoint subsets then define a tree order on this set of subsets.

We consider a binary tree system in which all the processes of depth d are synchronized. This is equivalent to defining a total order on the partition taken by grouping together processes of the same depth in the binary tree.

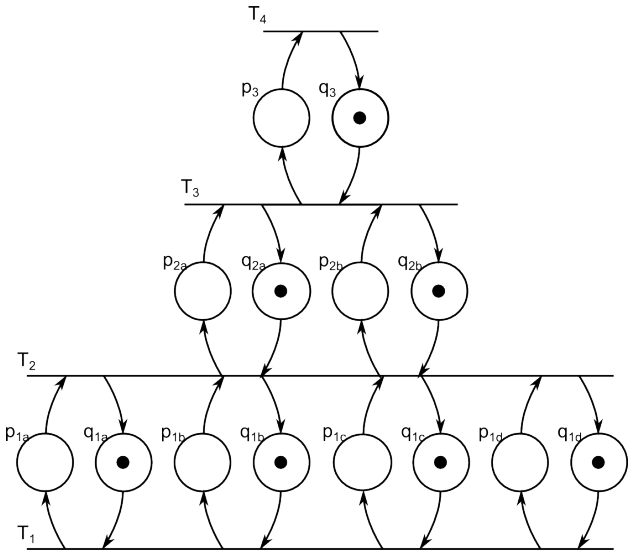


Figure 4.9: Petri net and initial marking for partially synchronous production binary tree for N=3.

Within this slightly more general framework the edge weight redistribution process still works in the same way.

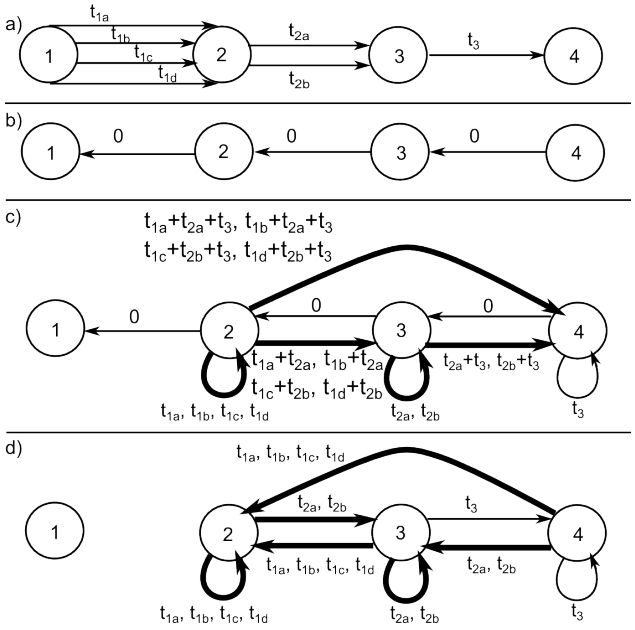


Figure 4.10: Graphs associated with A, B, A^*B and $\widehat{A}$ respectively for $N = 3$, bold edges represent sets of more than one edge whose weights are listed in the figure.

The adjacency matrix associated with the redistributed system is given by

$$\mathcal{A}_{i,j} = \begin{cases} 2^{i-1} & \text{for } j \leq i+1 \\ 0 & \text{otherwise} \end{cases} \tag{4.89}$$

and using

$$\min_{j=1}^{N} \sum_{i=1}^{N} \mathcal{A}_{i,j} \leq \Lambda \leq \max_{j=1}^{N} \sum_{i=1}^{N} \mathcal{A}_{i,j} \tag{4.90}$$

we have the inequality

$$2^{N-2} + 2^{N-1} = 3 \times 2^{N-2} \leq \Lambda \leq \sum_{j=1}^{N} 2^{j-1} = 2^N - 1 \tag{4.91}$$

so that

$$\lim_{N\to\infty} \frac{\lambda}{N} = \log 2 \tag{4.92}$$

4.4 Conclusion

4.4.1 Examples

Our simulations and bounds confirm that the asynchronous production tree has slightly lower throughput than the production line but that individual products are produced much faster. Enforced synchrony reduces throughput catastrophically but individual items are produced in a reasonable amount of time. We summarize our results in the following table which shows throughput and production time in the limit $N \to \infty$, so for very large production trees. $[\lambda_1 \approx 2.5] < [\lambda_2 \approx 3.5]$.

Production model	Throughput	Production time
Asynchronous Line	λ_1^{-1}	$\lambda_1 N$
Asynchronous tree	λ_2^{-1}	$\lambda_2 \log_2 N$
Partially synchronous tree	$(\log N)^{-1}$	$\log N \log_2 N$

4.4.2 Theory

For smaller values of Λ our bounds are not too sharp which is why we are unable to use them to approximate λ with much accuracy, however they still prove essential for proving the existence of the asymptote in λ as $N \to \infty$. For the partially synchronous case where $\Lambda \to \infty$ the bounds gives us an approximation of λ that converges in ratio as $N \to \infty$. The availability of such results is formalized in the corollary to our main theorem which we recall here.

Corollary Suppose that $[M_N(n)]_{n=1}^{\infty}$ is a sequence of i.i.d. componentwise exponential max-plus matrices parameterized by $N \in \mathbb{N}$.

- If $\max_N \Lambda_N = \overline{\Lambda}$ exists then $\max_N \lambda_N = \overline{\Lambda}$ exists.
- If on the other hand $\Lambda_N \to \infty$ then $\lim_{N\to\infty} \frac{\lambda_N}{\log \Lambda_N} = 1$
- Moreover if, as in all our examples, $M_N(n)$ can be realized as submatrix of $M_{N+K}(n)$ for all $N, K \geq 0$ then either $\lim_{N\to\infty} \Lambda_N$ exists and so does $\lim_{N\to\infty} \lambda_N$ or $\lim_{N\to\infty} \frac{\lambda_N}{\log \Lambda_N} = 1$

The ideas used in the proof of the bounds rest heavily on the maximally weighted path persecutive. In the proof of the lower bound we use a Markovian strategy to obtain a highly weighted path and in the proof of the lower bound we show exploit the special *Association* between different paths weights. These ideas could form the basis of a similar result for further classes of Max-plus matrices, indeed a similar upper bound can easily be obtained for *Componentwise Gaussian* matrices where we have

$$\lambda \asymp \sqrt{\log \Lambda} \tag{4.93}$$

The classical theory of graph eigenvalues gives us a useful toolbox for investigating the value of Λ in our examples which made its calculation fairly easy. In less structured systems it could still be obtained numerically with great accuracy much faster than any Max-plus exponent approximation.

4.4.3 Scope

The Max-plus formulation of a timed even graph will work for any such system with any sort of waiting time distribution. Our main theorem relied on a *Componentwise exponential* form which we obtained through a sort of combinatorial change of variables, this transformation will clearly not work for any timed event graph but similar techniques will no doubt be of use in other examples.

4.4.4 Further work

In further work we hope to generalize our result to a more general class of matrices, clearly the tail characteristics of the waiting time distributions play an important role here so something like the Pareto distribution whose tail weight can be parameterized would be particularly interesting. It would also be interesting to investigate what else can be said about the dynamics of a timed event graph from the maximally weighted path perspective, in particular the importance of the maximally weighted path's path, that is the sequence of vertices it visits. A statistical result would be especially desirable here.

Bibliography

[1] Bernd Heidergott, Geert Jan Olsder, Jacob van der Woude. (2005). Max Plus at Work. *Princeton Series in Applied Mathematics.* Vol 13

[2] Heidergott, Bernd F. (2007). Max-Plus Linear Stochastic Systems and Perturbation Analysis. *Springer.*

[3] Kolokoltsov, Vassili N., Maslov, Victor P. (1997). Idempotent Analysis and Its Applications *Springer.*

[4] David Speyer, Bernd Sturmfels. (2004). The tropical Grassmannian *Advances in Geometry.* Volume 4, No 3, Pages 389-411

[5] Itenberg, Ilia, Mikhalkin, Grigory, Shustin, Eugenii I. (2009). Tropical Algebraic Geometry. *Springer.*

[6] P. Butkovic. (2010). Max-linear Systems: Theory and Algorithms *Springer.*

[7] T Muranta .(1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE* Vol. 77, No. 4.

[8] C.A. Brackley, D. Broomhead, M. C. Romano and M. Thiel. (2011). A Max-Plus Model of Ribosome Dynamics During mRNA Translation. *arXiv:1105.3580v1*

Chapter 5

Smoothing non-smooth systems with the Moving Average Transformation

Abstract

We present a novel, systematic procedure for smoothing non-smooth dynamical systems. In particular we introduce the Moving Average Transformation which can be thought of as a change of variables which transforms discontinuous systems into *dynamically equivalent* continuous systems and continuous but non-differentiable systems into dynamically equivalent differentiable systems. This smoothing gives us a new way to compute the stability properties of a non-smooth systems and provides a new theoretical link between smooth and non-smooth systems. The dynamics and algebraic structure of systems obtained by transforming typical non-smooth systems are investigated.

5.1 Introduction

Discontinuities in non-smooth models of physical systems are approximations of highly non-linear smooth processes that take place on very short time scales. The

Newtonian impact of a ball bouncing on the ground is discontinuous in velocity but in reality the ball and ground are compliant, elastic bodies and discontinuities in velocity are physically impossible! Introducing compliance gives us a continuous but non-smooth system. The dynamics are non-smooth because there is only a force between the ball and ground when they are in contact, but this too is a false assumption. The electric fields of the atoms in the ball and ground are continually interacting in a smooth way it is just that the forces produced by these interactions are only significant during the short impacting phase of the dynamics.

Of course we approximate these fast processes by non-smooth discontinuities to obtain simpler, more manageable models of physical processes. Ironically the piecewise smooth nature of the resulting models can make them more difficult to study. For example non-differentiability makes it impossible to apply many standard numerical techniques to these systems [1, 2, 3, 4, 5]. For a discontinuous system like the bouncing ball the topology of the state space and any attractors it contains is broken up by jumps in the evolution so that an attractor will typically appear to be comprised of several disconnected parts, if we patch these together by connecting jump take offs and landings we arrive at a new topological space which can tell us much more about the system's dynamics than the topology of the original disconnected object.

There is therefore some motivation for transforming non-smooth systems (which as outlined are approximations of smooth systems) back into smooth systems. Several authors have tried a fairly direct approach, adding a small region where fast, smooth dynamics replace the non-smooth discontinuity in a process called regularization [5]. Whilst this doesn't necessarily mean introducing more layers of physically realistic modeling, since the extra components can be chosen to be as simple as possible, it still feels like a step in the wrong direction. The non-smooth discontinuities are a simplification and we should be able to exploit this trade off in realism by studying more mathematically appealing tractable systems.

In addition to these approximated physical processes there is a huge catalogue of non-smooth systems in control theory where digital switching between different

modes really should be thought of as being intrinsically non-smooth [6]. Abstract non-smooth systems are also of great interest mathematically as they can generically exhibit bifurcation structures which would be impossible or of high codimension in the space of smooth systems [7]. Although it is not natural to think of one of these systems as the limit of some smooth system it can still be advantageous to smooth them in some way as all the problems associated with the analysis of non-smooth physical systems are also present here.

In this chapter we introduce the Moving Average Transformation which can be thought of as a change of variables which transforms discontinuous systems into continuous systems and continuous but non-differentiable systems into differentiable systems. For a non-smooth system with evolution $\phi_t : X \to X$ the transformation is defined by

$$\Phi(x) = \int_{-1}^{0} \phi_\tau(x) d\tau \tag{5.1}$$

Since the evolution of the system is automatically incorporated into the transformation it is possible to systematically obtain smoothed systems using this technique. The definition of the transformation is reviewed in 5.1.2.

The transformation provides an explicit link between a non-smooth system and its smoothed *Dynamically Equivalent* counterpart. This enables us to better understand topological aspects of the dynamics and apply standard numerical techniques that rely on differentiability.

Crucially the technique is totally systematic, introduces no extraneous dynamics and provides a clear equivalence between the original and transformed (smoothed) system.

Although dynamical equivalence guaranties that all topological invariants other than continuity are preserved under the transformation and any standard smooth technique can be applied to the smoothed systems this chapter focuses on orbit stability. Small perturbations to an orbit in a non-smooth system evolve in a discontinuous way making stability analysis very complicated even in simple systems [3, 8]. In section 3 we show how the Saltation matrices which capture singular stability features of non-smooth discontinuities are naturally integrated into our smoothed

systems. Their discontinuous action is essentially spread out over a non-zero time interval of the flow.

In section 5.2 we demonstrate a novel way to calculate the stability of a periodic orbit in a non-smooth system. We start with a non-smooth system that contains a periodic orbit which we want to analyze. First we use the transformation to obtain a dynamically equivalent differentiable system. This smoothed system contains a periodic orbit corresponding to the orbit in the original system, we can then calculate the stability of this smoothed orbit in the usual smooth way, by integrating the derivative of the flow. The dynamical equivalence guarantees that the Lyupanov spectrum of orbits in the original and smoothed systems are the same. In section 5.4 we use the same idea, that the stability characteristics of the original and transformed systems are the same, to compute the Lyupanov exponent of a chaotic non-smooth system from a time-series recording. In principal this technique could be applied to experimental data.

This chapter is organized as follows. In the remainder of section 5.1 we introduce the moving average transformation and provide a simple example illustrating the sort of construction/analysis we hope to achieve using it. In section 5.2 we present a full and explicit example of applying the transformation twice to a discontinuous system. In section 5.3 we apply the transformation to the normal forms of some generic non-smooth discontinuities, this section provides most of the theory in the chapter and justifies our claim that the transformation smooths non-smooth systems, it also provides us with an atlas of possible local behaviors for typical smoothed systems. Finally in section 5.4 we smooth a more complicated system with a numerically implemented procedure based on a time-series reconstruction.

5.1.1 Simple example of smoothing

Before we formally introduce our method for smoothing we shall first by way of motivation, present a very simple example. Although the method here is quite ad hoc it illustrates the desired relationship between the original and smooth systems as

well as the necessary properties of the transformation between them.

Consider a unit mass attached to a spring with unit stiffness along with a wall where the mass undergoes Newtonian impacts. The wall is positioned at $x = 0$ where the spring is at its natural length . We assume that the spring is light and linear and that there is no friction so that away from the wall the dynamics are governed by the linear differential equation

$$\frac{d}{dt}\begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x \\ \dot{x} \end{pmatrix} \tag{5.2}$$

When the mass hits the wall there is an impact with restitution c so that whenever we reach the set $In = \{(x, \dot{x}) \in \mathbb{R}^2 : x = 0, \dot{x} < 0\}$ we instantaneously apply the map $R(x, \dot{x}) = (x, -c\dot{x})$ which maps In to $Out = \{(x, \dot{x}) \in \mathbb{R}^2 : x = 0, \dot{x} > 0\}$. Since we are mapped instantaneously from In to Out the state space of the system is given by $M = \{(x, \dot{x}) \in \mathbb{R}^2 : x \geq 0\}/\{(x, \dot{x}) \in \mathbb{R}^2 : x = 0, \dot{x} < 0\}$.

The system then evolves by flowing according to (5.2) until hitting In then mapping to Out and flowing according to (5.2) again... Define $\phi_t : M \mapsto M$ to be the time t evolution map that takes a point in the state space foreword in time t seconds.

We seek a smoothing transformation T which can be thought of as a change of variables with the property that the transformed system $\varphi_t = T \circ \phi_t \circ T^{-1}$ is smooth.

Claim The transformation $T : M \mapsto \{(y_1, y_2) \in \mathbb{R}^2\}$ defined in polar coordinates by

$$T(\theta, r) = (2\theta, rc^{\frac{\theta}{\pi}}) \tag{5.3}$$

provides an equivalence between our discontinuous system and the linear system evolved by the ODE

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\pi}\log c & 2 \\ -2 & \frac{1}{\pi}\log c \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{5.4}$$

defined on $\mathbb{R}^2$ equipped with the Euclidian metric.

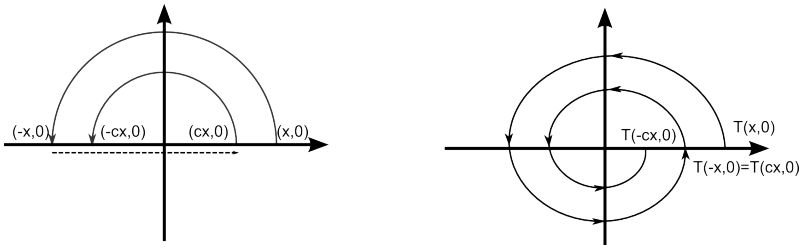


Figure 5.1: An orbit to the original system (Left) and its image under T (right).

Proof This rests quite heavily on our definition of equivalent! Clearly the two systems are not topologically equivalent since continuity is a topological invariant.

Definition If ϕ is the evolution of a non-smooth system which includes some jumps that are applied instantaneously on reaching a set Σ then we say that the transformation T provides a *Dynamical Equivalence* between ϕ and φ via $\varphi_t = T \circ \phi_t \circ T^{-1}$ if

- T is continuous
- T^{-1} exists and is continuous everywhere except $T(\Sigma)$
- $\frac{\delta T}{\delta x}$ exists and is non-zero everywhere

This is a relaxation of the usual definition for topological equivalence where T^{-1} is required to be continuous everywhere. Clearly our claim holds for this notion of equivalence.

Definition We define a *Cheat Metric* d on a discontinuous system to be a continuous metric which identifies jump take offs and landings so that $d(s, R(s)) = 0$ for all $s \in \Sigma$. To define d everywhere else we could just measure the length of the shortest path between two points but allow the path to jump over the discontinues in the same way as ϕ.

Viewed with the cheat metric all systems have continuous orbits, hence the cheat. This means that the original system with Euclidian metric is not topologically equivalent to the original system with cheat metric which in tern is topologically equivalent to the smoothed system with Euclidian metric.

The second condition for dynamical equivalence that T^{-1} exists and is continuous everywhere except $T(\Sigma)$ is equivalent to the condition that T^{-1} is continuous w.r.t a cheat metric. Therefore dynamically equivalent systems are topologically equivalent when viewed with a cheat metric.

The relationship between our original discontinuous system and the dynamically equivalent smooth system is quite subtle. If we equip M with the Euclidian metric then T^{-1} is discontinuous so the two systems are not topologically equivalent. However the evolutions are still interchangeable via $\varphi_\tau = T \circ \phi_\tau \circ T^{-1}$ so we can use the new system to describe the original. Since our new system is linear it is easy to solve and we can therefore obtain a neat expression for the evolution

$$\phi_t = T^{-1} \circ \frac{t}{\pi} \log c \begin{pmatrix} \sin 2t & \cos 2t \\ -\cos 2t & \sin 2t \end{pmatrix} \circ T \tag{5.5}$$

Complications caused by the discontinuity have all been captured in the transformation T. Since the new system is differentiable we can also calculate the stability of the fixed point at the origin which is determined by by the real part of (5.4)'s eigenvalue pair $\lambda_\pm = \frac{1}{\pi} \log c \pm 2i$. This agrees with the standard non-smooth analysis using the saltation matrix formulation [2].

We shall see that the moving average transformation has the same properties as T and can therefore be used to systematically obtain continuous systems on $\mathbb{R}^n$ dynamically equivalent to any discontinuous system we like. Moreover the transformation can be used to obtain differentiable systems topologically equivalent to continuous but non differentiable systems and so on so that each application of the transformation gives us an equivalent system one degree smoother.

5.1.2 The moving average transformation

Suppose that ϕ_t is the evolution operator of some possibly discontinuous or non differentiable dynamical system in a state space M which we assume is embedded in $\mathbb{R}^n$. The moving average transformation is then defined as a map $\Phi : \mathbb{R}^n \mapsto \mathbb{R}^n$ with

$$\Phi(x) = \int_{-1}^{0} \phi_\tau(x) d\tau \tag{5.6}$$

So that $\Phi(x)_i$ is the time averaged value of x's ith component over the last second of evolution. The averaging period of one second is totally arbitrary and from a theoretical point of view irrelevant, integrating back in time is also not important in fact anything of the form

$$\Upsilon(x) = \int_a^b \phi_\tau(x) d\tau \tag{5.7}$$

with $a < b$ will work in more or less the same way. We choose 1 for convenance and integrate backwards so that when applied to time-series we have something that makes sense from a signal processing point of view. For a discussion of applications to time-series and the averaging periods numerical importance see section 5.4.

In order for $\varphi_t = \Phi \circ \phi_t \circ \Phi^{-1}$ to be a dynamical equivalence between our original system and that obtained through the transformation we require that Φ is continuous and differentiable with non-zero derivative and an inverse that is continuous everywhere except the image of the jump set Σ.

The differentiability of Φ is a sufficient condition for our procedure to work (although we shall see that sliding systems which violate this condition can still be treated in this way). For a smooth flow the derivative is defined by considering the evolution of small perturbations to $\phi_{-1}(x)$ over one second. These perturbations obey the non-autonomous linear ODE

$$\frac{dz}{d\tau} = \frac{\delta^2 \phi_{\tau-1}(x)}{\delta\tau\delta x} z \tag{5.8}$$

Where the RHS is an $n \times n$ matrix of partial derivatives multiplying the n vector state. This equation has evolution operator $J(\tau)$ which is a τ dependent matrix. The derivative of Φ is then given by

$$\frac{\delta\Phi}{\delta x} = \int_0^1 J(\tau) d\tau J(1)^{-1} \tag{5.9}$$

If on the other hand the one second backwards time orbit of x encounters a non-smooth discontinuity like a jump or switch at $\phi_{-t^*}(x)$ with saltation S (see 5.2.1) then the derivative is given by

$$\frac{\delta\Phi}{\delta x} = \int_0^{1-t^*} J_1(\tau) d\tau [J_1(1-t^*) S J_2(t^*)]^{-1} + \int_0^{t^*} J_2(\tau) d\tau J_2(t^*)^{-1} \tag{5.10}$$

Where $J_1(\tau)$ is the evolution associated with perturbations about $\phi_{-1}(x)$ and $J_2(\tau)$ perturbations about $\phi_{-t^*}(x)$.

It is important to note that we do not need to know that saltation matrices to construct Φ, the construction and application of the transformation does not rely on any higher non-smooth theory. As shown in section 5.2 we can compute Φ by directly integrating along orbits, if we then differentiate our expression for Φ the saltation matrices naturally emerge in the analysis.

This is reviewed in section 5.3. That the derivative is non-zero is not guaranteed but can generically (in the topological sense) be remedied by taking a delay vector as in section 5.2 [9]. All that remains is to show that Φ^{-1} is continuous over discontinuities which is simple. If x and $F(x)$ are a preimage/image pair of points under a discontinuity in the system then $\phi_{-\epsilon}(x) = \phi_{-\epsilon}[F(x)]$ so the averages will agree.

Therefore we can apply Φ to a discontinuous system to obtain a dynamically equivalent continuous system. But we can actually Keep going. Since the transformation is able to encode information from the saltation matrices it is also able to transform continuous but non-differentiable systems (so ODEs with discontinuous RHSs) into differentiable systems (ODEs with continuous RHSs). Thus if we begin with a discontinuous system and apply the filter twice we obtain a continuous but non-differentiable system followed by a differentiable system and so on.

It is important to note that for every non-smooth event in the original system there will be two smoother but still non-smooth events in the transformed system one corresponding to the head of the averaging period crossing a discontinuity and another for the tail.

Definition The *Head Discontinuity* of a smoothed system corresponds to the head of the averaging period crossing the discontinuity in our original (untransformed) system. So that for example a switch at Σ in the original system gives rise to a head discontinuity in the smoothed system at $\Phi(\Sigma)$.

Definition The *Tail Discontinuity* of a smoothed system corresponds to the tail of the averaging period crossing the discontinuity in our original (untransformed)

system. So that for example a switch at Σ in the original system gives rise to a tail discontinuity in the smoothed system at $\Phi \circ \phi_1(\Sigma)$.

We shall see that for a jump or switch (5.3.1,2,3) the head and tail discontinuities are characteristically the same but that for slides and grazes (5.3.4,5,6,7) they are qualitatively different.

5.2 Explicit example

In this section we will explicitly obtain an algebraic form for the moving average transformation on a simple discontinuous system and apply it to obtain a dynamically equivalent continuous non-differentiable system. We then repeat the procedure to obtain an equivalent differentiable system. We shall then compute the stability of a periodic orbit in the differentiable system by integrating the derivative of the flow which we compare with a standard analysis of the discontinuous system.

Like in the motivating example our transformation will join jump take offs and landings in such a way that the new system is differentiable and dynamically equivalent to the original.

It should become clear through this section that the process of applying the filter and obtaining an induced flow is really intended as an exercise for computers rather than human beings! The various stages of solving differential equations, integrating solutions and inverting maps makes it difficult to find a non-trivial example which is still analytically tractable. The hope is that this example gives the reader a general idea of what happens and how. More complicated systems should be dealt with numerically either by the same direct method used here or with the time-series reconstruction approach outlined in section 5.4. Our system evolves according to

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 \\ ay \end{pmatrix} \tag{5.11}$$

along with the rule that at $\{(x, y) \in \mathbb{R}^2 : x = \pi\}$ we apply the map $R(x, y) = (x, by)$. Note that $\{(x, y) \in \mathbb{R}^2 : y = 0, x \in [0, \pi)\}$ is a periodic orbit independent of the

real parameters a and b, see Fig 2,a,i. Before we proceed with the moving average transformation we will review the calculation of saltation matrices and compute the stability of the system's periodic orbit with the standard non-smooth approach.

5.2.1 Calculation of saltation matrices

For a smooth flow $\dot{x} = F(x)$ the jacobian which governs the evolution of small perturbations to an orbit $\{\phi_t(x) : t \in \mathbb{R}_+\}$ is defined as the solution to the linear non-autonomous ODE

$$\dot{J}(t) = \frac{\delta^2 \phi_\tau(x')}{\delta x' \delta \tau}|_{\phi_t(x)} J(t) \tag{5.12}$$

which evolves continuously. Non-smooth discontinuities give rise to discontinuities in the evolution of these small perturbations and therefore the jacobian. We can capture these discontinuities in a linear way with a saltation matrix which is constructed for a jump as follows. Suppose that the flow is an ODE $\dot{x} = F(x)$ until reaching a set Σ where we instantaneously apply the map R then switch to the ODE $\dot{x} = G(x)$. For an orbit that hits Σ at a point s it suffices to assume that the flow is locally constant. Now take the following perturbations about s

- $a_1 = s - \epsilon F(s)$
- a_i to form a basis for the linearization of Σ at s

which after ϵ seconds are mapped by the flow to the following perturbations about $R(s) + \epsilon G(R(s))$

- $b_1 = -\epsilon G[R(s)]$
- $b_i = \frac{\delta R}{\delta x}|_s a_i + \epsilon G[R(s)]$

This transformation can be described uniquely by a linear map S, the saltation matrix. See Fig 2,b. For a switch or slide the analysis is exactly the same except that we set $R = I$.

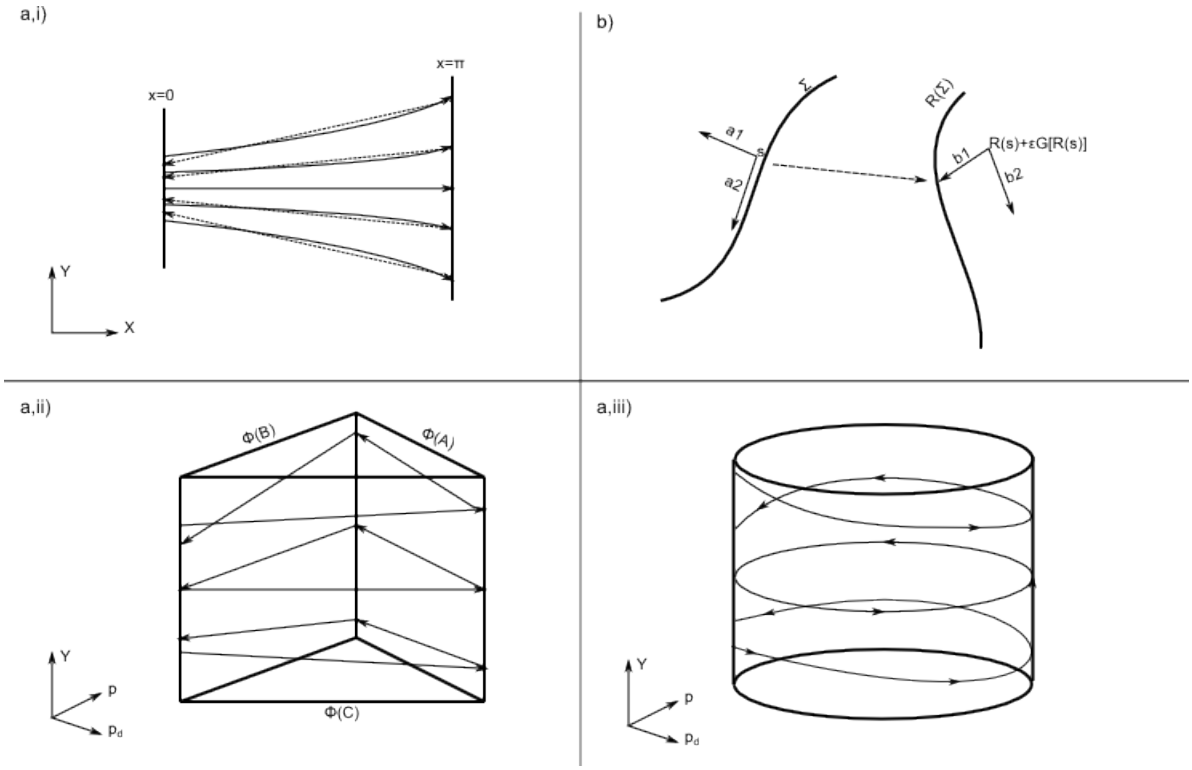


Figure 5.2: a,i) Orbits to original discontinuous system. ii) Continuous once transformed system. iii) Differentiable twice transformed system. b) Configuration of perturbations for saltation matrix calculation.

5.2.2 Stability of discontinuous system

The evolution of small perturbations along one period of our systems periodic orbit is given by $SJ(\pi)$ where $J(\pi)$ is the jacobian evaluated along the flow from $(0,0)$ to $(\pi,0)$ in the usual way and S is the saltation matrix associated with the jump at $(\pi,0)$. The jacobian is evolved by

$$\dot{J}(t) = \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix} J(t) \tag{5.13}$$

so that

$$J(\pi) = \begin{pmatrix} 1 & 0 \\ 0 & e^{a\pi} \end{pmatrix} \tag{5.14}$$

The saltation matrix is given by

$$S = \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix} \tag{5.15}$$

The stability of the orbit is determined by the eigenvalue $be^{a\pi}$ which corresponds to the Lyupanov exponent $a + \frac{\log b}{\pi}$.

5.2.3 Once transformed system

The moving average transformation

$$\Phi(x) = \int_{-1}^{0} \phi_t(x) dt \tag{5.16}$$

is piecewise smooth on two regions $A = \{0 \leq x \leq 1\}$ and $B = \{1 \leq x \leq \pi\}$. In A there is a jump in the 1-second backwards time flow so that

$$\Phi \begin{pmatrix} x \\ y \end{pmatrix} = \int_{-x}^{0} \begin{pmatrix} x+t \\ ye^{at} \end{pmatrix} dt + \int_{-1}^{-x} \begin{pmatrix} \pi + x + t \\ \frac{ye^{at}}{b} \end{pmatrix} \tag{5.17}$$

$$= \begin{pmatrix} x(1-\pi) + \pi - \frac{1}{2} \\ \frac{y}{a}[1 - \frac{e^{-a}}{b} + e^{-ax}(\frac{1}{b} - 1)] \end{pmatrix}$$

In B there is no jump in the 1-second backwards time flow so

$$\Phi \begin{pmatrix} x \\ y \end{pmatrix} = \int_{-1}^{0} \begin{pmatrix} x+t \\ ye^{at} \end{pmatrix} dt = \begin{pmatrix} x - \frac{1}{2} \\ \frac{y}{a}(1 - e^{-a}) \end{pmatrix} \tag{5.18}$$

If we now set $(p, q) = \Phi(x, y)$ and attempt to obtain the induced flow of (p, q) we run into some difficulties. In particular Φ is not injective so that the new system will require a hybrid formulation (that is an extra discrete variable to keep track of which branch of Φ^{-1} we are using). This situation is easily resolved however by making use of a delay vector. We add one delay with the map

$$D \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} & \varphi_{-1} \begin{pmatrix} p \\ q \end{pmatrix} \end{pmatrix} \tag{5.19}$$

So that the image of $D \circ \Phi$ is now a 2-dimensional manifold in $\mathbb{R}^4$. The transformation plus delay is an injection so that the new system admits a valid dynamical system

description. Conveniently the 2-dimensional manifold can be recovered after the following projection Π in the sense that Π^{-1} has a bijective branch which maps the image of $\Pi \circ D \circ \Phi$ to the image of $D \circ \Phi$. The projection is given by

$$\Pi \begin{pmatrix} p \\ q \\ p_D \\ q_D \end{pmatrix} = \begin{pmatrix} p \\ q \\ p_D \end{pmatrix} \tag{5.20}$$

where p_D and q_D are the delayed co-ordinates. This means that we only need to take a delay in the p variable to obtain a valid dynamical systems description and we can therefore represent everything in $\mathbb{R}^3$. Defining the composition $\Phi_D = \Pi \circ D \circ \Phi$ we have

$$\Phi_D \begin{pmatrix} x \\ y \end{pmatrix} = \begin{cases} \begin{pmatrix} x(1-\pi) + \pi - \frac{1}{2} \\ \frac{y}{a}[1 - \frac{e^{-1}}{b} + e^{-ax}(\frac{1}{b} - 1)] \\ \pi + x - \frac{3}{2} \end{pmatrix} & \text{in } A \\ \begin{pmatrix} x - \frac{1}{2} \\ \frac{y}{a}(1 - e^{-a}) \\ (x-1)(1-\pi) + \pi - \frac{1}{2} \end{pmatrix} & \text{in } B \\ \begin{pmatrix} x - \frac{1}{2} \\ \frac{y}{a}(1 - e^{-a}) \\ x - \frac{3}{2} \end{pmatrix} & \text{in } C \end{cases} \tag{5.21}$$

Where A is still $\{0 \leq x \leq 1\}$, B is now $\{1 \leq x \leq 2\}$ and $C = \{2 \leq x \leq \pi\}$. Differentiating (5.21) w.r.t. time using (5.11) then expressing in terms of (p, q, p_D) we obtain the discontinuous RHS for the ODE governing the now continuous transformed

system

$$\begin{pmatrix} \dot{p} \\ \dot{q} \\ \dot{p_D} \end{pmatrix} = \begin{cases} \begin{pmatrix} 1-\pi \\ \frac{aq(1-\frac{e^{-a}}{b})}{1-\frac{e^{-1}}{b}+e^{-a}\frac{p+\frac{1}{2}-\pi}{1-\pi}(\frac{1}{b}-1)} \\ 1 \end{pmatrix} & \text{in } \Phi_D(A) \\ \begin{pmatrix} 1 \\ aq \\ 1-\pi \end{pmatrix} & \text{in } \Phi_D(B) \\ \begin{pmatrix} 1 \\ aq \\ 1 \end{pmatrix} & \text{in } \Phi_D(C) \end{cases} \tag{5.22}$$

Where $\Phi_D(A), \Phi_D(B), \Phi_D(C)$ form the sides of an extruded triangle, see Fig 2,a,ii.

5.2.4 Twice transformed system

Successive applications of the moving average transformation smooths the system further, so a twice transformed discontinuous system will be differentiable. If the once transformed system's evolution operator is φ, as opposed to the original discontinuous systems operator ϕ then the transformation of the once transformed system is given by

$$\begin{pmatrix} P \\ Q \\ P_D \end{pmatrix} = \Phi \begin{pmatrix} p \\ q \\ p_D \end{pmatrix} = \int_{-1}^{0} \varphi_t \begin{pmatrix} p \\ q \\ p_D \end{pmatrix} dt = \underbrace{\Phi \circ \Pi \circ D \circ \Phi} \begin{pmatrix} x \\ y \end{pmatrix} \tag{5.23}$$

Now $\Pi \circ D$ commutes with Φ so the underbraced term equals $\Pi \circ D \circ \Phi^2$ and

$$\begin{pmatrix} P \\ Q \\ P_D \end{pmatrix} = \Pi \circ D \int_{-1}^{0} \int_{-1}^{0} \phi_{t_1+t_2} \begin{pmatrix} x \\ y \end{pmatrix} dt_1 dt_2 = \Pi \circ D \int_{-2}^{0} \phi_t \begin{pmatrix} x \\ y \end{pmatrix} h(t) dt \tag{5.24}$$

where

$$h(t) = \begin{cases} -t & \text{for } 0 \le t \le 1 \\ 2+t & \text{for } 1 \le t \le 2 \end{cases} \tag{5.25}$$

So that twice transforming the discontinuous system is equivalent to once applying the smoothing transformation

$$\Phi^2(x) = \Psi(x) = \int_{-2}^{0} \phi_t(x)h(t)dt \tag{5.26}$$

which when composed with $\Pi \circ D$ is piecewise smooth on 4 regions depending on the 3-second backwards time flow. As before depending on which of these regions we are in the backwards integrals are divided up into up to three separate parts all of which can easily be evaluated. In this way we obtain

$$\Psi_D \begin{pmatrix} x \\ y \end{pmatrix} = \begin{cases} \begin{pmatrix} \frac{-\pi}{2}x^2 + x + \pi - 1 \\ \frac{y}{a^2 b}[b - 2e^{-a} + e^{-2a} + e^{-ax}(1-b)(1+ax)] \\ \pi + x - 2 \end{pmatrix} & \text{in } A \\ \begin{pmatrix} \frac{\pi}{2}x^2 + x(1-2\pi) + 2\pi - 1 \\ \frac{y}{a^2 b}[b + e^{-2a} - 2be^{-a} + e^{-ax}[-1 + 2a + b - 2ab + x(ab-a)]] \\ \frac{-\pi}{2}x^2 + x(\pi+1) + \frac{\pi}{2} - 2 \end{pmatrix} & \text{in } B \\ \begin{pmatrix} x - 1 \\ \frac{y}{a^2}(1 + e^{-2a} - 2e^{-a}) \\ \frac{\pi}{2}x^2 + x(1-3\pi) + \frac{9\pi}{2} - 2 \end{pmatrix} & \text{in } C \\ \begin{pmatrix} x - 1 \\ \frac{y}{a^2}(1 + e^{-2a} - 2e^{-a}) \\ x - 2 \end{pmatrix} & \text{in } D \end{cases} \tag{5.27}$$

where $A = \{0 \leq x \leq 1\}$ $B = \{1 \leq x \leq 2\}$ $C = \{2 \leq x \leq 3\}$ and $D = \{3 \leq x \leq \pi\}$. Again we differentiate (5.27) using (5.11) then express in terms of (P, Q, P_D) to obtain the Continuous RHS for the ODE governing the now differentiable dynamics of the

twice transformed system.

$$\begin{pmatrix} \dot{P} \\ \dot{Q} \\ \dot{P}_D \end{pmatrix} = \begin{cases} \begin{pmatrix} -\pi P_D + \pi^2 - 2\pi + 1 \\ aQ + \frac{Qa^2 e^{-a\frac{P+P_D+3-\frac{5\pi}{2}}{2-\pi}} (\frac{P+P_D+3-\frac{5\pi}{2}}{2-\pi} - 2)(1-b)}{b+e^{-2a}-2be^{-a}+e^{-a\frac{P+P_D+3-\frac{5\pi}{2}}{2-\pi}}[-1+2a+b-2ab+\frac{P+P_D+3-\frac{5\pi}{2}}{2-\pi}(ab-a)]} \\ 1 \end{pmatrix} & \text{in } \Psi_D(A) \\ \begin{pmatrix} \frac{\pi}{2-\pi}(P+P_D+3-\frac{5\pi}{2}) + 1 - 2\pi \\ aQ + \frac{a^2 Q e^{-a\frac{P_D+P+3-\frac{5\pi}{2}}{2-\pi}} (b-1)(2-\frac{P_D+P+3-\frac{5\pi}{2}}{2-\pi})}{b+e^{-a}(1-2b)+e^{-a\frac{P_D+P+3-\frac{5\pi}{2}}{2-\pi}}[-1+b+(ab-a)(\frac{P_D+P+3-\frac{5\pi}{2}}{2-\pi}-2)]} \\ \frac{-\pi}{2-\pi}(P+P_D+3-\frac{5\pi}{2}) + 1 + \pi \end{pmatrix} & \text{in } \Psi_D(B) \\ \begin{pmatrix} 1 \\ aQ \\ \pi P + 1 - 2\pi \end{pmatrix} & \text{in } \Psi_D(C) \\ \begin{pmatrix} 1 \\ aQ \\ 1 \end{pmatrix} & \text{in } \Psi_D(D) \end{cases} \tag{5.28}$$

where $S = \Psi_D(A \cup B \cup C \cup D)$ forms a differentiable cylinder whose cross section is formed from three parabolas and one line segment, see Fig 2,a,iii.

5.2.5 Stability of differentiable system

Now we have a differentiable flow on a cylinder S we can calculate the stability of the periodic orbit in the usual way, by integrating the derivative. This could be a little tricky since the flow we are interested in lives on a 2-dimensional sub-manifold of $\mathbb{R}^3$ so we would normally have to compute the tangent space at each point on the orbit then integrate the derivative of the flow restricted to these tangencies. However the form of (5.28) gives us a perfectly good differentiable extension of the vector field onto $\mathbb{R}^3$ we can therefore integrate the whole derivative along the orbit and pick of the eigenvalues corresponding to perturbations in S two of which will exist since it is a 2-dimensional invariant set.

Perturbations about the point to the periodic orbit evolve according to the ODE

$$\begin{pmatrix} \delta\dot{P} \\ \delta\dot{Q} \\ \delta\dot{P}_D \end{pmatrix} = A(t) \begin{pmatrix} \delta P \\ \delta Q \\ \delta P_D \end{pmatrix} \tag{5.29}$$

where differentiating (5.28) gives

$$A(t) = \begin{cases} \begin{pmatrix} 0 & 0 & -\pi \\ 0 & a + \frac{a^2 b(b-1)e^{-at}t}{b-2e^{-a}+e^{-at}(at+1)(1-b)} & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{for } 0 \le t \le 1 \\ \begin{pmatrix} \frac{-\pi}{\pi-2} & 0 & \frac{-\pi}{\pi-2} \\ 0 & a + \frac{a^2 e^{-at}(b-1)(2-t)}{b+e^{-2a}-2be^{-a}+e^{-at}[-a+2a+b-2ab+t(ab-a)]} & 0 \\ \frac{\pi}{\pi-2} & 0 & \frac{\pi}{\pi-2} \end{pmatrix} & \text{for } 1 \le t \le 2 \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ \pi & 0 & 0 \end{pmatrix} & \text{for } 2 \le t \le 3 \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{for } 3 \le t \le \pi \end{cases} \tag{5.30}$$

Perturbations in P and P_D evolve in a piecewise autonomous way so we can separate and solve their evolution easily. The time-t map for this evolution is given by

$$J_{P,P_D} = \begin{cases} \begin{pmatrix} 1 & -\pi t \\ 0 & 1 \end{pmatrix} & \text{for } 0 \le t \le 1 \\ \begin{pmatrix} 1 - \frac{\pi t}{\pi-2} & \frac{-\pi t}{\pi-2} \\ \frac{\pi t}{\pi-2} & 1 + \frac{\pi t}{\pi-2} \end{pmatrix} \begin{pmatrix} 1 & -\pi \\ 0 & 1 \end{pmatrix} & \text{for } 1 \le t \le 2 \\ \begin{pmatrix} 1 & 0 \\ \pi t & 1 \end{pmatrix} \begin{pmatrix} \frac{-2}{\pi-2} & \frac{\pi}{\pi-2} \\ \frac{\pi}{\pi-2} & \frac{-\pi^2+2\pi-2}{\pi-2} \end{pmatrix} & \text{for } 2 \le t \le 3 \\ \begin{pmatrix} \frac{-2}{\pi-2} & \frac{\pi}{\pi-2} \\ \frac{-\pi}{\pi-2} & \frac{2\pi-2}{\pi-2} \end{pmatrix} & \text{for } 3 \le t \le \pi \end{cases} \tag{5.31}$$

The final matrix has an eigenvalue of 1 corresponding to the eigenvector $(1, 1)$ which

is tangent to the cylinder S at $\Psi_D(0,0)$. It can be verified that the evolution of this perturbation is always in the tangent space of S.

The evolution of perturbations in the Q co-ordinate can now be posed and solved as a 1-dimensional non-autonomous linear ODE the time t evolution of which is given by

$$J_Q = \begin{cases} \frac{e^{at}[b-2e^{-a}+e^{-2a}+e^{-ax}(1-b)(1+at)]}{1-2e^{-a}+e^{-2a}} & \text{for } 0 \leq t \leq 1 \\ \frac{e^{at}[b+e^{-2a}-2be^{-a}+e^{-ax}[-1+2a+b-2ab+x(ab-a)]]}{1-2e^{-a}+e^{-2a}} & \text{for } 1 \leq t \leq 2 \\ be^{at} & \text{for } 2 \leq t \leq \pi \end{cases} \tag{5.32}$$

which has the stability determining eigenvalue of $be^{a\pi}$ agreeing with the original system as claimed.

5.2.6 Remarks

We began with a discontinuous system with evolution operator ϕ and constructed a differentiable map Ψ_D and a differentiable system whose evolution operator we shall denote ψ. By construction the two systems are dynamically equivalent and the following commutation holds

$$\phi_t \begin{pmatrix} x \\ y \end{pmatrix} = \Psi_D^{-1} \circ \psi_t \circ \Psi_D \begin{pmatrix} x \\ y \end{pmatrix} \tag{5.33}$$

which means that via the smoothing transformation Ψ_D we can substitute one flow for the other.

Suppose we have some time series data obtained from the discontinuous system

$$F(t) = \langle \begin{pmatrix} a \\ b \end{pmatrix}, \phi_t \begin{pmatrix} x \\ y \end{pmatrix} \rangle \tag{5.34}$$

for $t \in [0, T]$ and we apply the smoothing filter to the series

$$G(t) = \int_{-2}^{0} F(t+\tau)h(\tau)d\tau \tag{5.35}$$

to obtain a new time series $\{G(t) : t \in [2, T]\}$. This time series is identical to the series we would record if instead of measuring the discontinuous system along some orbit

we measured the (smoothed) differentiable system along the corresponding smoothed orbit.

$$G(t) = \langle \begin{pmatrix} a \\ b \end{pmatrix}, \int_{-2}^{0} \phi_{t+\tau} \begin{pmatrix} x \\ y \end{pmatrix} h(\tau) d\tau \rangle = \langle \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \psi_t \Psi_D \begin{pmatrix} x \\ y \end{pmatrix} \rangle \tag{5.36}$$

In fact a stronger result holds. Generally we would not just measure $ax(t) + by(t)$ for our time-series but some possibly unknown smooth non-linear function of the state $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose that $\{F(t) = f \circ \phi_t(x) : t \in [0, T]\}$ is a time-series obtained from the discontinuous dynamical system then

$$G(t) = \int_{-1}^{0} F(t + \tau) d\tau = g \circ \varphi_t \circ \Phi_D(x) \tag{5.37}$$

for some continuous function g where Φ_D is the moving average transformation (applied just once) with delay and φ_t is the induced flow of the transformed system. To prove this we just set $g(p) = \Phi^f \circ \Phi_D^{-1}(p)$ where

$$\Phi^f \begin{pmatrix} x \\ y \end{pmatrix} = \int_{-1}^{0} f \circ \phi_t \begin{pmatrix} x \\ y \end{pmatrix} dt \tag{5.38}$$

Now Φ^f is continuous and Φ_D^{-1} is continuous everywhere except the transformation of the discontinuity surface $\Phi_D(\Sigma)$ where $\Phi_D^{-1}\Phi_D(\sigma)^- = \sigma$ and $\Phi_D^{-1}\Phi_D(\sigma)^+ = R(\sigma)$ but $\Phi^f(\sigma) = \Phi^f[R(\sigma)]$ so g is continuous everywhere.

Likewise if $\{F(t) = f \circ \phi_t(x) : t \in [0, T]\}$ is a time-series obtained from the discontinuous dynamical system then

$$G(t) = \int_{-2}^{0} F(t + \tau) h(\tau) d\tau = g \circ \psi_t \circ \Psi_D(x) \tag{5.39}$$

for some differentiable function g where Ψ_D is the double moving average transformation with delay and ψ is the induced flow of the transformed system.

These observations mean we can apply standard time series techniques that rely on differentiability to non-smooth systems. First record some possibly discontinuous time series, smooth it with a single or double application of the moving average filter then treat it as a differentiable recording from the differentiable smoothed system.

This give us a novel way to compute the stability of the discontinuous systems periodic orbit. First record a time series $\{F(t) = f \circ \phi_t(x) : t \in [0, T]\}$ with some arbitrary differentiable function f. Now apply the smoothing filter

$$G(t) = \int_{-2}^{0} F(t + \tau)h(\tau)dt \tag{5.40}$$

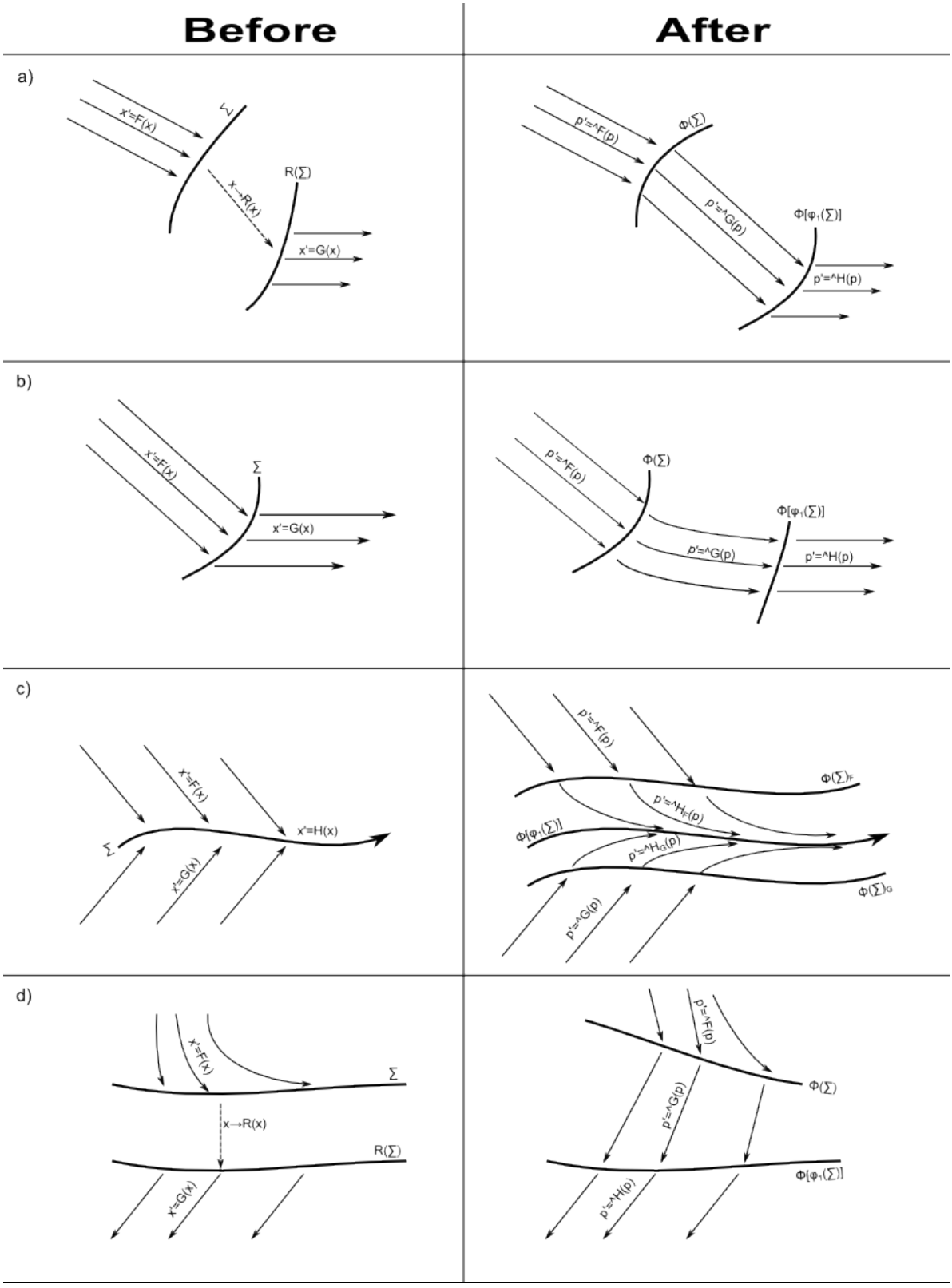
to obtain a new time series $\{G(t) = g \circ \psi_t \circ \Psi_D(x) : t \in [2, T]\}$ which we can think of as being obtained directly from the smoothed system by measuring with the differentiable function g. Standard numerical techniques enable us to calculate the lyupanov exponents of the smoothed system from this data [11] and since the stability of the discontinuous and smoothed systems are identical we have the lyupanov exponents of the original system.

On the flip-side we can also use these time-series reconstruction techniques to study the properties of the smoothed systems in the same way. Simulate the non-smooth system then smooth the data and treat it as coming directly from the smoothed system. In section 5.4 we use these ideas to investigate the dynamics of a more complex sliding oscillator.

5.3 Local analysis

Although it is difficult to obtain an explicit global description of the smoothed system $\varphi_t = \Phi \circ \phi_t \Phi^{-1}$ we can obtain local descriptions about some typical non-smooth discontinuities. By taking the piecewise linear normal form of a discontinuity we can formulate the smoothed system in the vicinity of the Φ image of the discontinuity and at the corresponding tail discontinuity. We shall see that the derivative of Φ captures the information in the saltation matrices and that this is then translated into the smoothed flow. The action of the filter is shown to transform discontinuous systems into continuous systems and continuous but non differentiable systems into differentiable systems further applications continue to smooth the system by one degree each time. Since the analysis is quite involved we summarize our results in the following table, see Fig 3.

Remark The following normal forms are not intended just as a series of simple examples but as an atlas of possible local behaviors for typical smoothed systems see 5.4.1.4.


Before
After
a)
x'=F(x)
x→R(x)
R(∑)
x'=G(x)
p'=^F(p)
Φ(∑)
p'=^G(p)
Φ[φ1(∑)]
p'=^H(p)
b)
x'=F(x)
∑
x'=G(x)
p'=^F(p)
Φ(∑)
Φ[φ1(∑)]
p'=^G(p)
p'=^H(p)
c)
x'=F(x)
x'=H(x)
∑
x'=G(x)
p'=^F(p)
Φ(∑)F
p'=^HF(p)
Φ[φ1(∑)]
p'=^HG(p)
Φ(∑)G
p'=^G(p)
d)
x'=F(x)
∑
x→R(x)
R(∑)
x'=G(x)
p'=^F(p)
Φ(∑)
p'=^G(p)
Φ[φ1(∑)]
p'=^H(p)

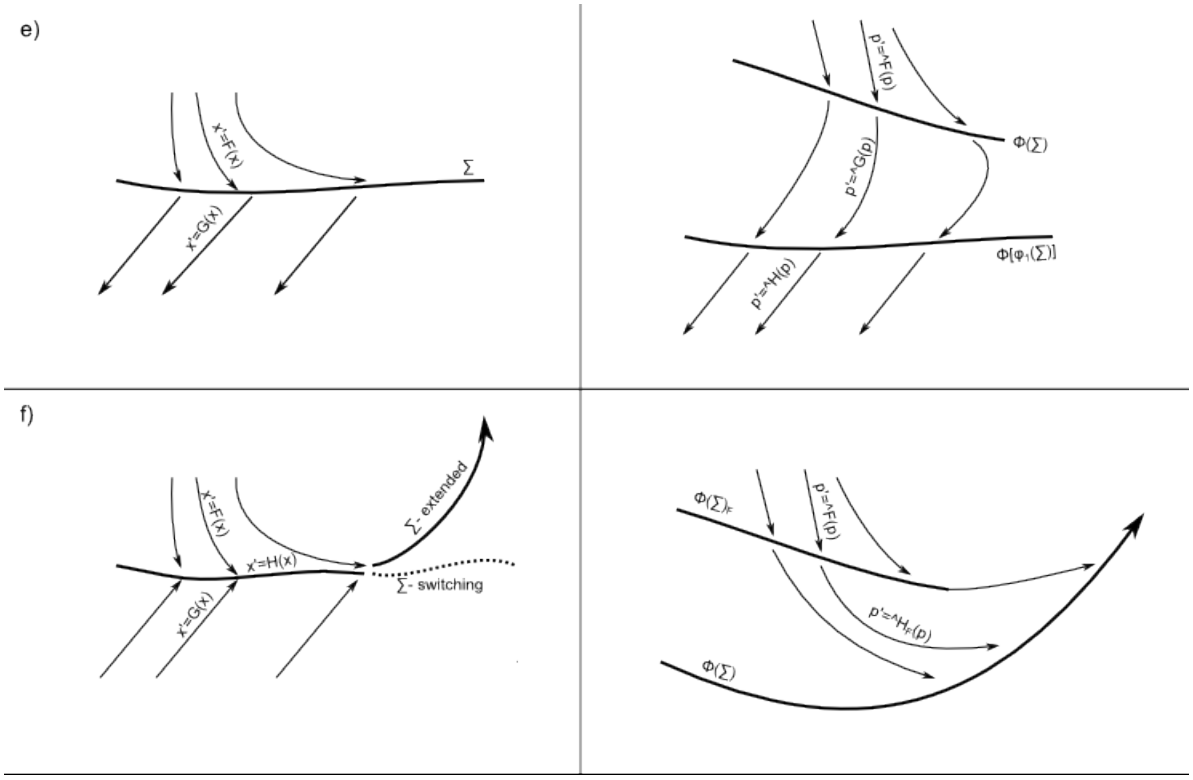


Figure 5.3: Non-smooth discontinuities and their smoothed counterparts. a) Discontinuous jump → two non-differentiable switches, b) Non-differentiable switch → two differentiable switches, c) Non-differentiable slide → differentiable switch followed by differentiable slide, d) Grazing discontinuous jump → grazing non-differentiable switch followed by non-differentiable switch, e) Grazing non-differentiable switch → grazing differentiable switch followed by differentiable switch, f) Grazing non-differentiable slide → grazing differentiable switch followed by differentiable slide.

Normal forms

We will consider 7 typical non-smooth discontinuities jumps, non-differentiable and k times differentiable switches, slides, jumping grazes, switching grazes and sliding grazes. In each case we formulate the simplest and most general non-degenerate form.

For the jump we assume that a differentiable vector field flows into a differentiable set where we apply a differentiable map, then after applying the map the flow is given by another differentiable vector field. We can therefore linearize everything in the vicinity of a point s on the discontinuity surface and its image after the jump R(s),

however this isn't the whole story since we also need information about the flow up to one second before the jump. In order to obtain an appropriate approximation (piecewise linear plus square root terms) of the transformed system at the image of the discontinuity $\Phi(s)$ we need to additionally know the one second backwards time position of the point $\phi_{-1}(s)$, the integrated jacobian J that maps perturbations about $\phi_{-1}(s)$ to perturbations about s and $\widetilde{J}$ the integral of the integrated jacobian along this orbit. It is worth noting that there is nothing to stop this one second backwards time flow containing any number of non-smooth discontinuities, everything important to the transformed system in the vicinity of the image of the discontinuity is contained in the point $\phi_{-1}(s)$ and pair of matrices $J, \widetilde{J}$ which can in principal be calculated or just assumed to be generic as is the case here.

Likewise when we analyze the tail discontinuity, the smoothed system in the vicinity of $\Phi \circ \phi_{+1}(s)$ the image of a point one second after the discontinuity surface, we will also need to know the point on the discontinuity surface we are flowing from s, its image under the jump map $R(s)$, the integrated jacobian J which maps perturbations about $R(s)$ to perturbations about s and its integral $\widetilde{J}$.

The remaining discontinuities are treated in much the same way with the normal form being outlined briefly at the beginning of each subsection along with some references for standard analysis and applications.

5.3.1 Jump

A jump is defined as follows. The flow is given by an ODE $\dot{x} = F(x)$ until reaching a set Σ where we instantaneously apply a map R, thereafter the flow is a second ODE $\dot{x} = G(x)$. Jumps are used to model Newtonian impacts [1] and arise as resets in some control systems [6] or when periodic variables are embedded as intervals rather than circles as in section 5.4. We will expand everything about a point $s \in \Sigma$ to obtain a piecewise linear description of the transformed system

$$\dot{p} = \frac{\delta x}{\delta t} \frac{\delta \Phi}{\delta x} \circ \Phi^{-1}(p) \tag{5.41}$$

about the point $\Phi(s)$.

Linearization of Φ before jump

We must differentiate Φ at s. By definition

$$\frac{\delta\Phi}{\delta x}|_s = \lim_{|h|\to 0} \frac{\Phi(s+h) - \Phi(h)}{|h|} = \lim_{|h|\to 0} \frac{\int_{-1}^{0} \phi_t(s+h) - \phi_t(s)dt}{|h|} \tag{5.42}$$

which can be expressed in terms of the jacobians integrated from $\phi_{-1}(s)$ so that the derivative is given by

$$[\int_0^1 J(t)dt]J(1)^{-1} = \widetilde{J}J^{-1} \tag{5.43}$$

where J maps perturbations about $\phi_{-1}(s)$ to perturbations about s and $\widetilde{J}$ maps perturbations about $\phi_{-1}(s)$ to the integral of the perturbation evolving over one second. The linearization is given by

$$\Phi(x) = \Phi(s) + \widetilde{J}J^{-1}(x - s) \tag{5.44}$$

Linearization of Φ after jump

The Linearization of Φ at $R(s)$ is a little more complicated. We might try to use the formula $\widetilde{J}(SJ)^{-1}$ for the derivative since S^{-1} maps perturbations about $R(s)$ to perturbations about s. However this is not the whole story since

$$\Phi(x) = \Phi(s) + \widetilde{J}(SJ)^{-1}[x - R(s)] \tag{5.45}$$

does not include any contribution to the integral from after the jump and instead contains a spurious contribution from the wrong side of the discontinuity surface. Figure 4 shows the configuration of s, $R(s)$, and x with the various components of $\Delta\Phi$.

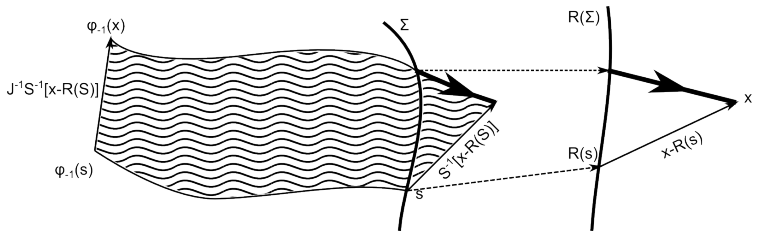


Figure 5.4: Hatched region represents jacobian approximation $\widetilde{J}(SJ)^{-1}[x - R(s)]$ we need to remove the spurious contribution from the wrong side of Σ and add the contribution to Φ from after the jump (both bold).

We can correct for these contributions by assuming that the flow is locally constant at s and $R(s)$. The contribution from after the jump is given by

$$\int_{-t^*}^{0}(x+G[R(s)]t)dt \tag{5.46}$$

where t^* the length of time since the perturbation made the jump is given by

$$t^* = \frac{\langle x - R(s), n_2\rangle}{\langle G[R(s)], n_2\rangle} \tag{5.47}$$

where n_2 is normal to $R(\Sigma)$ at $R(s)$. The spurious contribution from the wrong side of the jump is given by

$$\int_{t^*}^{0}(s+S^{-1}[x-R(s)]+F(s)t)dt \tag{5.48}$$

where t^* the length of time that the perturbation is taken to be on the wrong side of the jump given by

$$t^* = \frac{\langle S^{-1}[x-R(s)], n_1\rangle}{\langle F(s), n_1\rangle} \tag{5.49}$$

where n_1 is normal to Σ at s. It follows from the definition of the saltation matrix that the two expressions for t^* are the same. Finally the linearization is given by

$$\Phi(x) = \Phi[R(s)] + \widetilde{J}(SJ)^{-1}[x-R(s)] + t^*[R(s)-s] \tag{5.50}$$

To invert the transformation we need to express t^* in terms of $\Phi(x)$. First we have

$$x = R(s) + SJ\widetilde{J}^{-1}[\Phi(x) - \Phi[R(s)] - t^*(R[s]-s)] \tag{5.51}$$

and

$$t^* = \frac{\langle J\widetilde{J}^{-1}[\Phi(x) - \Phi[R(s)] - t^*(R[s]-s)], n_1\rangle}{\langle F(s), n_1\rangle} \tag{5.52}$$

so that

$$t^* = \frac{\langle J\widetilde{J}^{-1}(\Phi(x) - \Phi[R(s)]), n_1\rangle}{\langle F(s) + J\widetilde{J}^{-1}[R(s)-s], n_1\rangle} \tag{5.53}$$

Transformed system

To obtain a piecewise linear description of the transformed system we need to be able to express $\dot{p}$ as a function of p. This would first require the following expansion

$$\dot{p} = [F(s) + \frac{\delta F}{\delta x}|_s(x-s)][\frac{\delta \Phi}{\delta x}|_s + \frac{\delta \frac{\delta \Phi}{\delta x}|_{x'}}{\delta x'}|_s(x-s)] \tag{5.54}$$

which we could then use the linearization of Φ to put in terms of p rather than x. However we do not need to compute the complicated second derivative of Φ. Instead we use the following

$$\dot{p} = \lim_{h\to 0} \frac{p(t+h) - p(t)}{h} = \lim_{h\to 0} \frac{\int_{-1}^{0} \phi_{\tau+h}(x) - \phi_{\tau}(x) dt}{h} = x - \phi_{-1}(x) \tag{5.55}$$

Which can easily be linearized so that before the jump we have

$$\dot{p} = x - \phi_{-1}(s) - J^{-1}(x - s) \tag{5.56}$$

and afterwards we have

$$\dot{p} = x - \phi_{-1}(s) - J^{-1}S^{-1}[x - R(s)] \tag{5.57}$$

Thus our transformed system first flows according to the ODE $\dot{p} = \widehat{F}(p)$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (I - J^{-1})J\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.58}$$

Until we reach the set $\Phi(\Sigma)$ whose linearization about $\Phi(s)$ is given by

$$\{p : \langle J\widetilde{J}^{-1}[p - \Phi(s)], n_1\rangle = 0\} \tag{5.59}$$

where we instantaneously switch to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = R(s) - \phi_{-1}(s) + (SJ - I)\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.60}$$
$$+ \frac{\langle J\widetilde{J}^{-1}(\Phi(x) - \Phi[R(s)]), n_1\rangle}{\langle F(s) + J\widetilde{J}^{-1}[R(s) - s], n_1\rangle}(I - SJ)\widetilde{J}^{-1}[R(s) - s]$$

Tail discontinuity

There will be a second non-smooth event associated with the jump in the smoothed system. One second after crossing the jump the tail of the averaging period will cross the jump resulting in a second switch in the smooth system on the switching surface $\Phi \circ \phi_1(\Sigma)$. The analysis here is more or less the same as before. To see this consider the backwards evolution of the jumping system where we flow according to the ODE

$\dot{(}x) = -G(x)$ until reaching the set $R(\Sigma)$ where we instantaneously apply the map R^{-1} them flow according to the second ODE $\dot{x} = -F(x)$. The transformation of this system in the vicinity of the jump will be an ODE of the form $\dot{x} = H(x)$ which defines the forewords system at the second switch by $\dot{x} = -H(x)$.

Since $\dot{x} = H(x)$ will be of the same form as the transformed system in 3.1.3 the reversed system at the tail discontinuity $\dot{x} = -H(x)$ will also be a non-differentiable switch consisting of a differentiable vector field flowing into a differentiable manifold at which point we switch to another differentiable vector field.

5.3.2 Non-differentiable switch

A switch is defined as follows. The dynamics are governed by an ODE $\dot{x} = F(x)$ which flows into a set Σ where we instantaneously switch to a second ODE $\dot{x} = G(x)$ and do not switch back. Switches are used to model discontinuous changes of force in mechanics [1] and discontinuous changes to control input in closed loop control systems [6]. We will linearize everything about a point $s \in \Sigma$ to obtain a piecewise linear description of the transformed system about the point $\Phi(s)$. The switch can therefore be thought of as a special case of the jump with $R = I$ the identity so can therefore go straight to the transformed system.

The flow is given by the ODE $\dot{p} = \widehat{F}(p)$ whose linearization at $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (J - I)\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.61}$$

Until we reach the set $\Phi(\Sigma)$ whose linearization at $\Phi(s)$ is given by

$$\{p : \langle J\widetilde{J}^{-1}[p - \Phi(s)], n_1\rangle = 0\} \tag{5.62}$$

where we switch to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization at $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (SJ - I)\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.63}$$

For $p \in \Phi(\Sigma)$ we have $SJ\widetilde{J}^{-1}[p - \Phi(s)] = J\widetilde{J}^{-1}[p - \Phi(s)]$ since $J\widetilde{J}^{-1}[p - \Phi(s)]$ will lie on the original switching surface where the saltation matrix acts as the identity.

Therefore the two expressions (5.61) and (5.63) agree on the switching surface and the system is differentiable.

Likewise the tail discontinuity is a differentiable switch, which we can prove in the same way we treated the jump in 5.3.1.4.

5.3.3 k-times differentiable switch

A k-times differentiable switch is a switch as before except that the first k derivatives of the vector fields F and G agree on the switching manifold Σ. In order to analyze this system we must introduce some non-linear terms in the analysis which we do using the exact saltation function $\mathcal{S}$. The inverse of this function maps a point in the space governed by G to the point that it's backwards time pre-image in the space governed by F would have reached had it not switched vector fields on reaching Σ. Thus is ϕ^F and ϕ^G are the evolution operators associated with the two flows we have

$$\mathcal{S}^{-1}(x) = \phi^F_{t(x)} \circ \phi^G_{-t(x)}(x) \tag{5.64}$$

where $t(x)$ is the least t such that $\phi^G_{-t(x)}(x) = \Sigma(x) \in \Sigma$, which exists and is smooth away from any grazing points. Note that the derivative of $\mathcal{S}$ at a point $s \in \Sigma$ is exactly the saltation matrix S at s, which for a differentiable switch equals the identity. We will treat only the head discontinuity since as in the non-differentiable switch the head and tail are essentially the same.

Dynamics before the switch

The transformation of the system before Σ is governed by the ODE

$$\dot{p} = \widehat{F}[\Phi_F^{-1}(p)] \tag{5.65}$$

where $\widehat{F}(x) = x - \phi_1^F(x)$ and $\Phi_F(x) = \int_{-1}^{0} \phi_\tau^F(x) d\tau$ is smooth with smooth inverse.

After the switch

The dynamics are now governed by the ODE

$$\dot{p} = \widehat{G}[\Phi_G^{-1}(p)] \tag{5.66}$$

where

$$\widehat{G}(x) = x - \phi^F_{-1} \circ \mathcal{S}^{-1}(x) \tag{5.67}$$

and

$$\Phi_G(x) = \Phi_F[\mathcal{S}^{-1}(x)] + \underbrace{\int_0^{t(x)} \phi^G_\tau[\Sigma(x)] - \phi^F_\tau[\Sigma(x)] d\tau} \tag{5.68}$$

Transformed system

The transformed system is a $k+1$ times differentiable switch. To prove this we will show that the first $k+1$ derivatives of $\widehat{F}$ and $\widehat{G}$ agree on Σ as well as the first $k+1$ of Φ_F and Φ_G.

Since the original system is k-times differentiable and therefore has $k+1$ times differentiable orbits we can write

$$x = \Sigma(x) + a_1 t(x) + ... + a_{k+1} t(x)^{k+1} + b_{k+2} t(x)^{k+1} + ... \tag{5.69}$$

and

$$\mathcal{S}^{-1}(x) = \Sigma(x) + a_1 t(x) + ... + a_{k+1} t(x)^{k+1} + c_{k+2} t(x)^{k+1} + ... \tag{5.70}$$

so that

$$\mathcal{S}^{-1}(x) = x + (c_{k+2} - b_{k+2}) t(x)^{k+2} + ... \tag{5.71}$$

Therefore on Σ where $t(x) = 0$ we have

$$\tfrac{\delta \mathcal{S}^{-1}(x)}{\delta x} = I \quad \tfrac{\delta^m \mathcal{S}^{-1}(x)}{\delta x^m} = 0, m = 2, 3, ..., k+1 \quad \tfrac{\delta^{k+2} \mathcal{S}^{-1}(x)}{\delta x^{k+2}} \neq 0 \tag{5.72}$$

Which proves that the first $k+1$ derivatives of $\widehat{F}$ and $\widehat{G}$ agree on Σ. All that remains is to show that the first $k+1$ derivatives of the underbraced term in (5.68) are zero expanding the functions of τ we have

$$\int_0^{t(x)} (b_{k+2} - c_{k+2}) \tau^{k+2} + ... d\tau = (b_{k+2} - c_{k+2}) \frac{\tau^{k+3}}{k+3} \tag{5.73}$$

the first $k+2$ derivatives of which will be zero on Σ.

5.3.4 Slide

A slide is defined as follows. We have a switching surface Σ where the dynamics are governed by an ODE $\dot{x} = F(x)$ on one side of Σ and another ODE $\dot{x} = G(x)$ on the other side of Σ. For $s \in \Sigma$ to be a sliding point we require $\langle F(s), n\rangle$ and $\langle G(s), n\rangle$ to have opposite signs where n is normal to Σ at s. So if we employed something like an Euler scheme then when we reach (and generically overshoot) Σ evolving according to $x_{n+1} = x_n + \tau F(x_n)$ we are forced to switch back and forth between the two RHSs trapped in a small neighborhood of Σ which collapses onto Σ as $\tau \to 0$.

In order to properly define the solution of the non smooth ODE we use the Filippov formalism in which we consider the RHS of our differential equation as being a well defined set valued function obtained by taking the convex hull of all possible values of the discontinuous vector field at each point

$$\dot{x} \in \begin{cases} \{F(x)\} & \text{for } \langle x, n\rangle > 0 \\ \{aF(x) + (1-a)G(x) : a \in [0,1]\} & \text{for } \langle x, n\rangle = 0 \\ \{G(x)\} & \text{for } \langle x, n\rangle < 0 \end{cases} \tag{5.74}$$

where n is normal to Σ at the closest point in Σ to x. It can then be shown that the solution to this *Differential Inclusion* is well defined and satisfies the following ODE on the switching surface [10]

$$\dot{x} = H(x) = \frac{F^{\perp}(x)|G^{\parallel}(x)| + G^{\perp}(x)|F^{\parallel}(x)|}{|G^{\parallel}(x)| + |F^{\parallel}(x)|} \tag{5.75}$$

Where $F^{\parallel}$ and $F^{\perp}$ are F's components parallel and perpendicular to n respectively - likewise for $G^{\parallel}, G^{\perp}$. Slides are used to model systems with static friction, see section 5.4 or [1, 2, 8].

Example

There are some interesting complications here so we will first look at a simple example. Our system flows according to the non-smooth ODE

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 \\ \begin{cases} -1 & \text{for } y > 0 \\ 1 & \text{for } y < 0 \end{cases} \end{pmatrix} \tag{5.76}$$

which defines the flow on $\{y = 0\}$ as $\dot{x} = 1$.

Clearly this system is not time-reversible on $y = 0$ where Φ will have infinitely branches. However suppose we take any $(x', y') \in \mathbb{R}^2$ and compute its forward time orbit $\{\phi_t(x', y') : t \in \mathbb{R}_+\}$ then calculate Φ from this orbit for each $(x, y) \in \{\phi_t(x', y') : t \in [1, \infty)\}$ calling this new quantity

$$\begin{pmatrix} p \\ q \end{pmatrix} = \Phi_{(x',y')} \begin{pmatrix} x \\ y \end{pmatrix} \tag{5.77}$$

Now we can differentiate the above to obtain

$$\begin{pmatrix} \dot{p} \\ \dot{q} \end{pmatrix} = \lim_{h\to\infty} \frac{\Phi_{(x',y')} \begin{pmatrix} x + \delta(x) \\ y + \delta(y) \end{pmatrix} - \Phi_{(x',y')} \begin{pmatrix} 0 \\ y \end{pmatrix}}{h} \tag{5.78}$$

which is fine since $(x + \delta(x), y + \delta(y)) \in \{\phi_t(x', y') : t \in [1, \infty)\}$. So we obtain the differentiable system

$$\begin{pmatrix} \dot{p} \\ \dot{q} \end{pmatrix} = \begin{pmatrix} 1 \\ \begin{cases} -1 & \text{for } q > \frac{1}{2} \\ \frac{-q^{\frac{1}{2}}}{2} & \text{for } \frac{1}{2} \geq q > 0 \\ 0 & \text{for } q = 0 \\ \frac{(-q)^{\frac{1}{2}}}{2} & \text{for } 0 > q \geq \frac{-1}{2} \\ 1 & \frac{-1}{2} > q \end{cases} \end{pmatrix} \tag{5.79}$$

Because of the choice of branch the switching surface is fattened by the transformation to a higher dimensional object. In this case from a line to a strip. Those points on the strip but not on the line $\{q = 0\}$ correspond to points that have joined the sliding surface less than one second ago, see Fig 3,c. In the discontinuities we have looked at so far there are two non-smooth events associated with them in the transformed system, one corresponding to x crossing the discontinuity and another corresponding to $\phi_{-1}(x)$ crossing the discontinuity - the Head and the Tail of our averaged period. In the jump and switch the two events where essentially the same and therefore didn't require any extra analysis but for the slide the two are clearly very different. In this example the first non-smooth event at $q = \frac{\pm 1}{2}$ is a differentiable switch but the second

at $q = 0$ is quite different with parabolic solutions converging differentially and in finite time onto the switching surface. To treat the general case of the switch we must look at these two discontinuities separately.

Head discontinuity

The flow is given by an ODE $\dot{x} = F(x)$ until reaching a set Σ where we instantaneously switch to an ODE $\dot{x} = H(x)$ as in (5.75). We will linearize everything about a point $s \in \Sigma$ to obtain a piecewise linear description of the transformed system.

Before reaching the slide the linearization of Φ is as before

$$\Phi(x) = \Phi(s) + \widetilde{J}J^{-1}(x - s) \tag{5.80}$$

As is the flow

$$\dot{p} = s - \phi_{-1}(s) + (I - J^{-1})(x - s) \tag{5.81}$$

So that the transformed system flows according the ODE $\dot{p} = \widehat{F}$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (I - J^{-1})J\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.82}$$

After reaching the slide things are a little more complex. To represent points on the sliding surface we use (σ, t) co-ordinates. Here $\sigma \in \Sigma$ is the current position of the point and $t \ll 1$ is the branching index for the backwards flow, it tells us for how long the backwards orbit from σ stays on the sliding surface. See Fig 5,a.

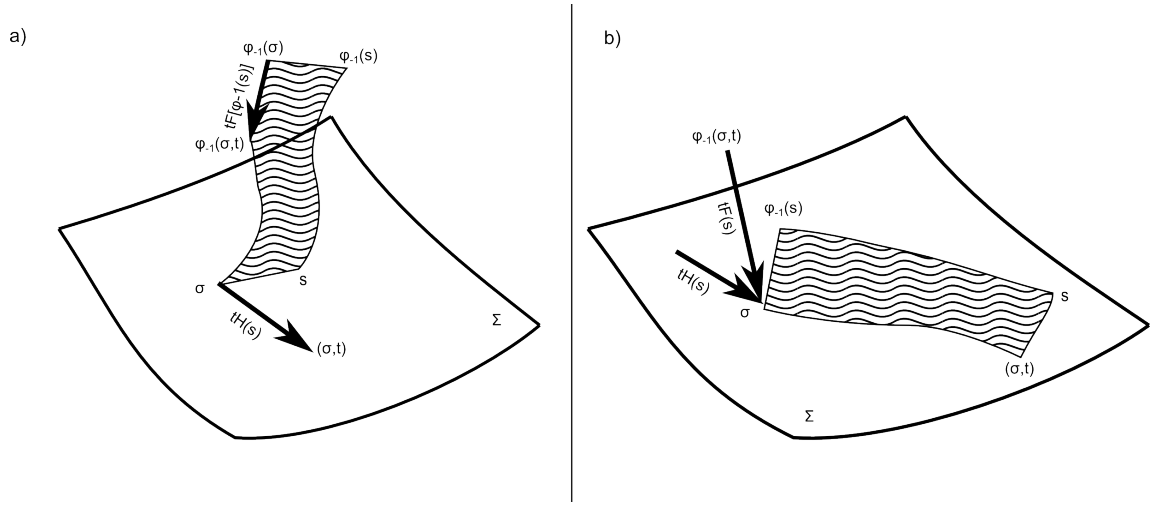


Figure 5.5: Configuration of components in $\Delta\Phi$; jacobian approximation (hatched) and correction terms (bold) for the slide at a) the head and b) tail discontinuities.

The linearized transformation is given by

$$\Phi(\sigma,t) = \Phi(s)+\underbrace{\widetilde{J}J^{-1}[\phi_{-t}(\sigma,t)-s]}+\int_0^t \phi_{-t}(\sigma,t)+\tau H(s)d\tau-\int_0^t \phi_{1+t}(\sigma,t)+\tau F[\phi_{-1}(s)]d\tau \tag{5.83}$$

$$= \Phi(s) + \widetilde{J}J^{-1}[\sigma - tH(s) - s] + t[s - \phi_{-1}(s)]$$

where the underbraced term is the derivative of Φ applied at the point where (σ,t) first joins the sliding surface and the two integrals corrects for the short period of sliding. To invert the transformation we note that $\sigma - tH(s) - s$ is tangent to Σ at s so that if n is the corresponding normal vector then taking the dot product with $(J\widetilde{J}^{-1})^T n$ gives

$$t = \frac{\langle \Phi(\sigma,t) - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle s - \phi_{-1}(s), (J\widetilde{J}^{-1})^T n\rangle} \tag{5.84}$$

and

$$\sigma = s + J\widetilde{J}^{-1}(\Phi(\sigma,t) - \Phi(s) - t[s - \phi_{-1}(s)]) + tH(s) \tag{5.85}$$

The flow is given by

$$\dot{p} = \sigma - \phi_{-1}(\sigma,t) = \sigma - (\underbrace{\phi_{-1}(s) + J^{-1}[\sigma - tH(s) - s]}) - tF[\phi_{-1}(s)] \tag{5.86}$$

Where the underbraced term is the one second backwards time image of the point where (σ,t) first joins the sliding surface and the final term corrects for the fact that

(σ, t)'s one second backwards time image is actually the $(1-t)$ second backwards time image of this point. So that the transformed system flows according to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (J - I)\widetilde{J}^{-1}(p - \Phi(s)) \tag{5.87}$$
$$+\frac{\langle p - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle s - \phi_{-1}(s), (J\widetilde{J}^{-1})^T n\rangle}(I - J^{-1})[H(s) - F(s)]$$

Expressions (5.82) and (5.87) agree on the switching surface where $t = 0$ so the switch is differentiable.

Tail discontinuity

The flow is given by an ODE $\dot{x} = F(x)$ until reaching a set Σ close to a point s where we instantaneously switch to an ODE $\dot{x} = H(x)$ as in (5.75). After 1 second orbits reach the vicinity of $\phi_1(s)$. From the example we do not expect a linear description of the transformed system to adequately illustrate the dynamics so we will include some higher order terms. Also since the system $\dot{x} = H(x)$ lives on an $n-1$ dimensional manifold we must work with some arbitrary smooth continuation of H when computing the jacobians - of course the exact nature of the continuation is unimportant - its existence just means we don't need to be too careful about what we say!

Again we will work in (σ, t) co-ordinates with $\sigma \in \Sigma$ the current position of the point only this time the branch index $t \ll 1$ tells us that the backwards time orbit of (σ, t) stays on the sliding surface for $(1-t)$ seconds, see Fig 5,b. Including some second order terms in t the transformation about $\phi_1(s)$ is given by

$$\Phi(\sigma, t) = \Phi[\phi_1(s)] + \underbrace{\widetilde{J}J^{-1}(\sigma - \phi_1(s))} + \int_{-t}^{0} \phi_{1-t}(\sigma) + \tau F(s) d\tau - \int_{-t}^{0} \phi_{1-t}(\sigma) + \tau H(s) d\tau \tag{5.88}$$
$$= \Phi[\phi_1(s)] + \widetilde{J}J^{-1}(\sigma - \phi_1(s)) + \frac{t^2}{2}[H(s) - F(s)]$$

Where the underbraced term is the derivative of Φ restricted to the backwards flow branch that stays on the sliding surface for all time and the two integrals correct for

the t seconds that we are not on the sliding surface. We invert the transformation in the same way so that

$$\frac{t^2}{2} = \frac{\langle \Phi(\sigma,t) - \Phi[\phi_1(s)], (J\widetilde{J}^{-1})^T n\rangle}{\langle H(s) - F(s), (J\widetilde{J}^{-1})^T n\rangle} \tag{5.89}$$

where n is normal to Σ at $\phi_1(s)$.

$$\sigma = \phi_1(s) + J\widetilde{J}^{-1}(\Phi(\sigma,t) - \Phi[\phi_1(s)]) - \frac{t^2}{2}[H(s) - F(s)]) \tag{5.90}$$

The flow is given by

$$\dot{p} = \sigma - \phi_{-1}(\sigma,t) = \sigma - (\underbrace{s + J^{-1}[\sigma - \phi_1(s)]}) + t[H(s) - F(s)] \tag{5.91}$$

where the underbraced term is the one second backwards time image of σ restricted to the branch where we stay on the sliding surface for all time and the final term corrects for the t seconds that we are not on the sliding surface. So that the transformed system flows according to the ODE $\dot{p} = \widehat{H}(p)$ whose expansion about $\Phi[\phi_1(s)]$ is given by

$$\dot{p} = a + B(p - \Phi[\phi_1(s)]) + c\sqrt{\langle p - \Phi[\phi_1(s)], d\rangle} \tag{5.92}$$

where

$$\begin{aligned} a + B(p - \Phi[\phi_1(s)]) &= \phi_1(s) - s + (J - I)\widetilde{J}^{-1}(p - \Phi[\phi_1(s)]) \\ &+ \frac{\langle \Phi(\sigma,t) - \Phi[\phi_1(s)], (J\widetilde{J}^{-1})^T n\rangle}{\langle H(s) - F(s), (J\widetilde{J}^{-1})^T n\rangle}(J - I)\widetilde{J}^{-1}[H(s) - F(s)] \end{aligned} \tag{5.93}$$

and

$$c\sqrt{\langle p - \Phi[\phi_1(s)], d\rangle} = [H(s) - F(s)]\sqrt{\frac{\langle \Phi(\sigma,t) - \Phi[\phi_1(s)], (J\widetilde{J}^{-1})^T n\rangle}{\langle H(s) - F(s), (J\widetilde{J}^{-1})^T n\rangle}} \tag{5.94}$$

Thus as in the simple example orbits converge differentially and in finite time to the transformed sliding surface $\Phi(\Sigma) = \{\langle J\widetilde{J}^{-1}(p - \Phi[\phi_1(s)]), n\rangle = 0\}$ and the distance decays like $(t - c)^2$ due to the square root in the equation for the flow.

5.3.5 Grazing a jump

A jump graze is defined as follows. The dynamics are governed by the ODE $\dot{x} = F(x)$ which flows into a set Σ where we instantaneously apply a map R, thereafter the

dynamics are governed by a second ODE $\dot{x} = G(x)$. For $s \in \Sigma$ to be a grazing point we require $\frac{\delta d(x)}{\delta t}|_s = 0$ where $d(x)$ measures the distance from x to the switching surface Σ. Thus arbitrarily small perturbations to s result in large immediate changes to the evolution. Grazes are typically whenever we have a jump, switch or slide and are responsible for singular stability properties and novel bifurcations in non-smooth systems [2, 3, 7].

If we attempt to construct a Saltation Matrix to capture the effects of the jump at a grazing point we run into trouble. Recall that for a saltation matrix S associated with a switch we require

- $SF(s) = SG(s)$ where F and G are the value of the flows vector field before and after the switch.
- $S(a_i) = \frac{\delta R}{\delta x}|_s a_i$ where the a_i $i = 1, 2, ..., n-1$ provide a basis for the linearization of Σ at s.

we do not know what to choose for the vector field after the jump F or G? This doesn't really matter tho as either choice leads to disaster. Taking $H(s) = (F/G)(s)$ gives

$$S = \begin{pmatrix} & & \cdots & \\ H(s) & \frac{\delta R}{\delta x}|_s a_1 & \cdots & \frac{\delta R}{\delta x}|_s a_{n-1} \\ & & \cdots & \end{pmatrix} \begin{pmatrix} & & \cdots & \\ F(s) & a_1 & \cdots & a_{n-1} \\ & & \cdots & \end{pmatrix}^{-1} \tag{5.95}$$

and the second matrix is non-invertible since F must be tangential to Σ at s and therefore in the span of the a_i. We can however still obtain a piecewise linear description of the transformed system - we are just unable to make use of the saltation formulation in any of the analysis.

In the standard analysis of grazing orbits we make use of some non-linear terms which come into play as the singularity at the graze elevates them to the linear level through a characteristic square root action. Likewise we will use the time since hitting the discontinuity as a co-ordinate which in the switching and sliding case appears in some non-linear terms. In this jumping case the smoothed system doesn't contain any

square root terms but still contains a non-differentiable graze which has the required singular properties.

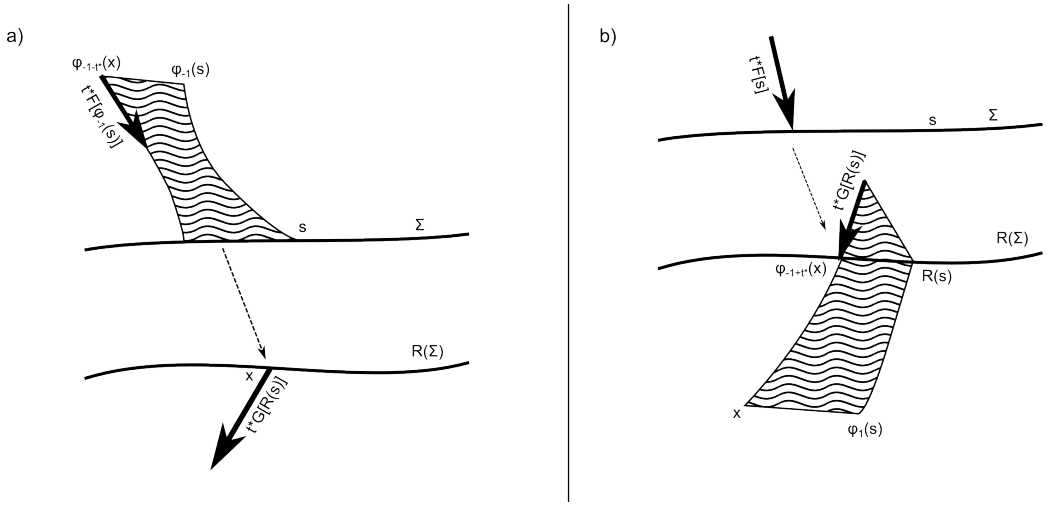


Figure 5.6: Configuration of components in $\Delta\Phi$; jacobian approximation (hatched) and correction terms (bold) for the jumping graze at a) the head and b) tail discontinuities.

Before the jump the analysis is exactly the same as with the non-grazing case so that the transformed system flows according to the ODE $\dot{p} = \widehat{F}(p)$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (I - J^{-1})\widetilde{J^{-1}}[p - \Phi(s)] \tag{5.96}$$

until we reach the set $\Phi(\Sigma)$ whose linearization about $\Phi(s)$ is given by

$$\{p : \langle \widetilde{J^{-1}}[p - \Phi(s)], n_1 \rangle = 0\} \tag{5.97}$$

where n_1 is normal to Σ at s. Note that $\Phi(s)$ is a grazing point of the new system since $\Phi(s) + \epsilon\dot{p}|_{\Phi(s)}$ lies in $\Phi(\Sigma)$.

$$\widetilde{J^{-1}}[\Phi(s) + \epsilon\dot{p}|_{\Phi(s)} - \Phi(s)] = \widetilde{J^{-1}}\epsilon\dot{p}|_{\Phi(s)} \tag{5.98}$$

$$\widetilde{J^{-1}}\epsilon[s - \phi_{-1}(s)] = J\epsilon F[\phi_{-1}(s)] = \epsilon F(s)$$

and by definition $\langle F(s), n_1 \rangle = 0$.

After the jump the linearization of Φ about $R(s)$ is given by

$$\Phi(x) = \Phi(s) + \underbrace{\widetilde{J}J^{-1}(\frac{\delta R}{\delta x}|_s)^{-1}[x - t^*G[R(s)] - R(s)]} \tag{5.99}$$

$$+ \int_0^{t^*} \phi_{-t^*}(x) + \tau G[R(s)]d\tau - \int_0^{t^*} \phi_{1+t^*}(x) + \tau F[\phi_{-1}(s)]d\tau$$

$$= \Phi(s) + \widetilde{J}J^{-1}(\frac{\delta R}{\delta x}|_s)^{-1}[x - t^*G[R(s)] - R(s)] + t^*[R(s) - \phi_{-1}(s)]$$

Where the underbraced term is the derivative of Φ at the point where x hits Σ and the integrals correct for the t^* seconds of flow after the jump, see Fig 6,a, where

$$t^* = \frac{\langle x - R(s), n_2\rangle}{\langle G[R(s)], n_2\rangle} \tag{5.100}$$

with n_2 normal to $R(\Sigma)$ at $R(s)$. We invert the transformation as in 5.3.3.2 so that

$$t^* = \frac{\langle \Phi(x) - \Phi(s), (J\widetilde{J}^{-1})^T n_1\rangle}{\langle R(s) - \phi_1(s), (J\widetilde{J}^{-1})^T n_1\rangle} \tag{5.101}$$

where n_1 is normal to Σ at s.

$$x = R(s) + \frac{\delta R}{\delta x}|_s J\widetilde{J}^{-1}[\Phi(x) - \Phi(s) - t^*(R[s] - \phi_{-1}[s])] + t^*G[R(s)] \tag{5.102}$$

The flow is given by

$$\dot{p} = x - \phi_{-1}(x) = x - \underbrace{(\phi_{-1}(s) + J^{-1}\frac{\delta R}{\delta x}|_s^{-1}[x - t^*G[R(s)] - R(s)]} + t^*F[\phi_{-1}(s)]) \tag{5.103}$$

Where the underbraced term is the one second backwards time image of the point where x's orbit hits Σ and the final term corrects for the fact that x's one second backwards time image is this point's $(1 - t^*)$ second backwards time image. So that the transformed systems flows according to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = R(s) - \phi_1(s) + (I - J^{-1}\frac{\delta R}{\delta x}|_s^{-1})\frac{\delta R}{\delta x}|_s \tag{5.104}$$

$$\times \underbrace{J\widetilde{J}^{-1}[p - \Phi(s) - \frac{\langle p - \Phi(s), (J\widetilde{J}^{-1})^T n_1\rangle}{\langle R(s) - \phi_1(s), (J\widetilde{J}^{-1})^T n_1\rangle}(R[s] - \phi_{-1}[s])]}$$

$$+ \frac{\langle p - \Phi(s), (J\widetilde{J}^{-1})^T n_1\rangle}{\langle R(s) - \phi_1(s), (J\widetilde{J}^{-1})^T n_1\rangle}(F[\phi_{-1}(s)] - G[R(s)])$$

where the underbraced term is tangent to Σ at s so there are no problems in applying R's derivative or its inverse.

Tail discontinuity

Again the tail discontinuity requires a slightly different analysis. If we consider the backwards system as with the jump and switch we have an orbit which grazes the image of the jumping surface after going over the jump which is not qualitatively the same as the foreword system. The flow is given by an ODE $\dot{x} = F(x)$ until we reach a set Σ where we instantaneously apply a map R then flow according to a second ODE $\dot{x} = G(x)$. For $s \in \Sigma$ to be a grazing point we require $\frac{\delta d(x)}{\delta t}|_s = 0$ where $d(x)$ measures the distance from x to the switching surface Σ. We will linearize everything about the point $\phi_1(s)$.

The switching surface for this second discontinuity will be given by $\Phi[\phi_1(\Sigma)]$ whose linearization at $\Phi[\phi_1(s)]$ is given by

$$\{p : \langle \widetilde{J}^{-1}(p - \Phi[\phi_1(s)), n_2\rangle = 0\} \tag{5.105}$$

where J maps perturbations about $R(s)$ to perturbations about $\phi_1(s)$, $\widetilde{J}$ is the integral along these perturbations evolution as before and n is normal to $R(\Sigma)$ at $R(s)$. After flowing for more than one second from the jump the transformed system flows according to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization about $\Phi[\phi_1(s)]$ is given by

$$\dot{p} = \phi_1(s) - R(s) + (1 - J^{-1})J\widetilde{J}^{-1}(p - \Phi[\phi_1(s)]) \tag{5.106}$$

Before crossing the switching surface the linearization of the transformation is given by

$$\Phi(x) = \Phi[\phi_1(s)] + \underbrace{\widetilde{J}J^{-1}[x - \phi_1(s)]} \tag{5.107}$$

$$\begin{aligned}\lim_{\epsilon\to 0}\int_{-t^*}^{0} \phi_{-1+t^*-\epsilon}(x) + \tau F(s)d\tau - \int_{-t^*}^{0} \phi_{-1+t^*}(x) + \tau G[R(s)]d\tau \\ = \Phi[\phi_1(s)] + \widetilde{J}J^{-1}[x - \phi_1(s)] + t^*[s - R(s)]\end{aligned} \tag{5.108}$$

Where the underbraced term is the derivative of Φ assuming no jump has taken palace and the two integrals account for the jump that occurred $(1 - t^*)$ seconds ago, see Fig 6,b, where

$$t^* = \frac{\langle J^{-1}[x - \phi_1(s)], n_2\rangle}{\langle G[R(s)], n_2\rangle} \tag{5.109}$$

We invert the transformation directly

$$x = \phi_1(s) + J\widetilde{J}^{-1}[\Phi(x) - \Phi[\phi_1(s)] - t^*(s - R[s])] \tag{5.110}$$

and

$$t^* = \frac{\langle \widetilde{J}^{-1}[\Phi(x) - \Phi[\phi_1(s)] - t^*(s - R[s])], n_2\rangle}{\langle G[R(s)], n_2\rangle} \tag{5.111}$$

so that

$$t^* = \frac{\langle \widetilde{J}^{-1}(\Phi(x) - \Phi[\phi_1(s)]), n_2\rangle}{\langle G[R(s)], n_2\rangle(1 + \langle [s - R(s)], n_2\rangle)} \tag{5.112}$$

The flow is given by

$$\dot{p} = x - \phi_{-1}(x) = x - \underbrace{[s + (\frac{\delta R}{\delta x}|_s)^{-1}(J^{-1}[x - \phi_1(s)] + t^* G[R(s)])]}_{} + t^* F(s) \tag{5.113}$$

Where the underbraced term is the point where x's orbit hit Σ and the final term corrects for the t^* seconds of flow before this. So that before reaching the discontinuity the transformed system flows according to the ODE $\dot{p} = \widehat{F}(p)$ whose linearization about $\Phi[\phi_1(s)]$ is given by

$$\dot{p} = \phi_1(s) - s + [J - (\frac{\delta R}{\delta x}|_s)^{-1}]\widetilde{J}^{-1}(p - \Phi[\phi_1(s)]) \tag{5.114}$$

$$+\frac{\langle \widetilde{J}^{-1}(p - \Phi[\phi_1(s)]), n_2\rangle}{\langle G[R(s)], n_2\rangle(1 + \langle [s - R(s)], n_2\rangle)}(F(s) - [J - (\frac{\delta R}{\delta x}|_s)^{-1}]\widetilde{J}^{-1}[s - R(s)] - (\frac{\delta R}{\delta x}|_s)^{-1} G[R(s)])$$

Note that there are no grazes associated with this non-differentiable switch. After one second

$$\langle \widetilde{J}^{-1}[\phi_1(s) - R(s)], n_2\rangle = \langle G[R(s)], n_2\rangle \neq 0 \tag{5.115}$$

and before $[\phi_1(s) - s]$ bares no special relationship with $\widetilde{J}^{-1}$ so should also yield a non-zero product.

5.3.6 Grazing a switch

A switching graze is defined as follows. The flow is given by an ODE $\dot{x} = F(x)$ until reaching a set Σ where we instantaneously switch to a second ODE $\dot{x} = G(x)$. For $s \in \Sigma$ to be a grazing point we require $\frac{\delta d(x)}{\delta t}|_s = 0$ where $d(x)$ measures the distance from x to the switching surface Σ.

Grazing a Switch is therefore a special case of Grazing a Jump with $R = I$. Before the Switch everything is the same however we cannot use the same analysis after the switch. The linearized transformation is given by

$$\Phi(x) = \Phi(s) + \widetilde{J}J^{-1}[x - t^*G(s) - s] + t^*[s - \phi_{-1}(s)] \tag{5.116}$$

but $s - \phi_{-1}(s) = \widetilde{J}J^{-1}F(s)$ so

$$\Phi(x) = \Phi(s) + \widetilde{J}J^{-1}\underbrace{[x - t^*G(s) - s + t^*F(s)]} \tag{5.117}$$

and the underbraced term is tangent to Σ at s which is $n-1$ dimensional. The linearized transformation is therefore not invertible. We expand Φ with some second order terms in t^*

$$\Phi(x) = \Phi(s) + \underbrace{\widetilde{J}J^{-1}[x - t^*G(s) - s]} \tag{5.118}$$

$$+ \int_0^{t^*} \phi_{-t^*}(x) + \tau G(s) d\tau - \int_0^{t^*} \phi_{1+t^*}(x) + \tau F[\phi_{-1}(s)] d\tau$$

$$= \Phi(s) + \overbrace{\widetilde{J}J^{-1}[x - t^*G(s) - s] + t^*[s - \phi_{-1}(s)]} + \frac{(t^*)^2}{2}[G(s) - F(\phi_{-1}[s])] \tag{5.119}$$

Where the underbraced term is the derivative of Φ applied at the point where x's orbit hits Σ and the two integrals correct for the t^* seconds of flow after the switch. To invert the transformation we note that the overbraced term is tangent to $\Phi(\Sigma)$ at $\Phi(s)$ so that taking the dot product with $(J\widetilde{J}^{-1})^T n$ where n is normal to Σ at s gives

$$\frac{(t^*)^2}{2} = \frac{\langle \Phi(x) - \Phi(s), (J\widetilde{J}^{-1})^T n \rangle}{\langle G(s) - F[\phi_{-1}(s)], (J\widetilde{J}^{-1})^T n \rangle} \tag{5.120}$$

and

$$x = s + t^*G(s) + J\widetilde{J}^{-1}[\Phi(x) - \Phi(s) - t^*(s - \phi_{-1}[s]) - \frac{(t^*)^2}{2}(G[s] - F[\phi_{-1}(s)])] \tag{5.121}$$

The flow is given by

$$\dot{p} = x - \phi_{-1}(x) = x - (\underbrace{\phi_{-1} + J^{-1}[x - t^*G(s) - s]} + t^*F[\phi_{-1}(s)]) \tag{5.122}$$

where the underbraced term is the one second backwards time image of the point at which x's orbit hits Σ and the final term corrects for the fact that this $\phi_{-1}(x)$ is

this point's $(1-t^*)$ backwards time image. So that the transformed system flows according the ODE $\dot{p} = \widehat{G}(p)$ whose expansion about $\Phi(s)$ is given by

$$\dot{p} = a + B[p - \Phi(s)] + c\sqrt{\langle p - \Phi(s), d\rangle} \tag{5.123}$$

where

$$a + B[p - \Phi(s)] = s - \phi_{-1}(s) + (J - I)\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.124}$$
$$-\frac{\langle p - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle G(s) - F[\phi_{-1}(s)], (J\widetilde{J}^{-1})^T n\rangle}(J - I)\widetilde{J}^{-1}[G(s) - F(\phi_1[s])]$$

and

$$c\sqrt{\langle p - \Phi(s), d\rangle} = [G(s) - F(s)]\sqrt{2\frac{\langle p - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle G(s) - F[\phi_{-1}(s)], (J\widetilde{J}^{-1})^T n\rangle}} \tag{5.125}$$

Note that (5.212) agrees with the flow before the switch on $t = 0$ so the graze is differentiable.

Tail discontinuity

The analysis follows exactly as in the case of grazing a jump only we set $R = I$. The transformed systems flow is an ODE ODE $\dot{p} = \widehat{F}(p)$ whose linearization about $\Phi[\phi(s)]$ is given by

$$\dot{p} = \phi_1(s) - s + (I - J^{-1})J\widetilde{J}^{-1}(p - \Phi[\phi_1(s)]) \tag{5.126}$$
$$+ \underbrace{\frac{\langle \widetilde{J}^{-1}(p - \Phi[\phi_1(s)]), n\rangle}{\langle G(s), n\rangle}}[F(s) - G(s)]$$

until reaching the set $\Phi[\phi_1(\Sigma)]$ whose linearization at $\Phi[\phi_1(s)]$ is given by

$$\{p : \langle \widetilde{J}^{-1}[p - \Phi \circ \phi_1(s)], n\rangle = 0\} \tag{5.127}$$

Where the flow switches to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization about $\Phi[\phi(s)]$ is given by

$$\dot{p} = \phi_1(s) - s + (1 - J^{-1})J\widetilde{J}^{-1}(p - \Phi[\phi_1(s)]) \tag{5.128}$$

For a point p on the new switching surface $\widetilde{J}^{-1}(p - \Phi[\phi_1(s)])$ lies on the original switching surface so the second underbraced term is zero and the expressions in (5.216) and (5.218) agree so that the switch is differentiable.

5.3.7 Grazing a slide

A Sliding graze is defined as follows. We have a switching surface Σ where the flow is given by an ODE $\dot{x} = F(x)$ on one side of Σ and another ODE $\dot{x} = G(x)$ on the other side of Σ. For $s \in \Sigma$ to be a sliding point we require $\langle F(s), n\rangle$ and $\langle G(s), n\rangle$ to have opposite signs where n is normal to Σ at s. As in (5.75) we use the Fillipov formalism to define the ODE $\dot{x} = H(x)$ on Σ. For $s \in \Sigma$ to be a grazing point we require $\frac{\delta d(x)}{\delta t}|_s = 0$ where $d(x)$ measures the distance from x to the switching surface Σ.

Therefore the set of grazing points form the boundary of the sliding region. If a point joins the sliding surface and flows along it to some grazing point s' it will then leave the sliding surface since at s' the sliding condition is violated and provided the derivative of F at s' is non-degenerate the point will then flow according to $\dot{x} = F(x)$ moving away from the switching surface. We can extend our sliding surface Σ to include all forwards time orbits from the set of grazing points and extend H the vector field on the sliding surface to this extended set by using F restricted to the new manifold, see Fig 7. Within this framework we can analyze the sliding graze in more or less the same way we treated the basic slide. We will expand everything about a sliding graze point s to obtain an approximation of the transformed system.

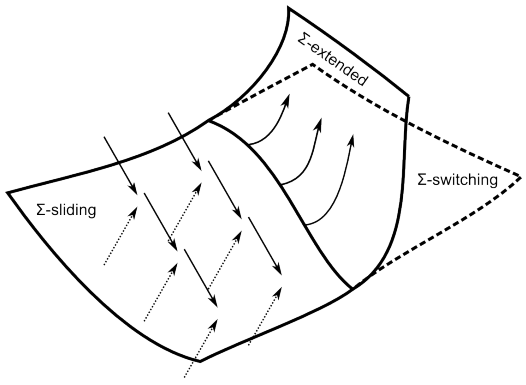


Figure 5.7: Extended sliding manifold Σ

Before reaching the discontinuity everything is as before so that the transformed system flows according to the ODE $\dot{p} = \widehat{F}(p)$ whose linearization about $\Phi(s)$ is given

by

$$\dot{p} = s - \phi_{-1}(s) + (I - J^{-1})J\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.129}$$

After reaching the discontinuity surface we are stuck to the extended $(n-1)$ dimensional manifold Σ. Again we use (σ, t) co-ordinates where $\sigma \in \Sigma$ is the current position of the point and $t \ll 1$ is the branching index which tells us for how long its backwards orbit stays on the sliding surface. Including second order terms in t the transformation at s is given by

$$\Phi(\sigma,t) = \Phi(s) + \underbrace{\widetilde{J}J^{-1}[\phi_{-t}(\sigma,t) - s]} + \int_0^t \phi_{-t}(\sigma,t) + \tau H(s)d\tau - \int_0^t \phi_{1+t}(\sigma,t) + \tau F[\phi_{-1}(s)]d\tau \tag{5.130}$$

$$= \Phi(s) + \widetilde{J}J^{-1}[\sigma - tH(s) - s] + t[s - \phi_{-1}(s)] + \frac{t^2}{2}(H(s) - F[\phi_{-1}(s)])$$

where the underbraced term is the derivative of Φ applied at the point where (σ, t) first joins the sliding surface and the two integrals correct for the t seconds of sliding. Note that since $F(s) = H(s)$ the terms that are linear in t cancel

$$\widetilde{J}J^{-1}F(s) = s - \phi_{-1}(s) \tag{5.131}$$

hence the need for the second order terms. We invert the transformation as before so that

$$\frac{t^2}{2} = \frac{\langle \Phi(\sigma,t) - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle F(s) - F[\phi_{-1}(s)], (J\widetilde{J}^{-1})^T n\rangle} \tag{5.132}$$

where n is normal to Σ at s.

$$\sigma = s + J\widetilde{J}^{-1}[\Phi(\sigma,t) - \Phi(s) - \frac{t^2}{2}(F(s) - F[\phi_1(s)])] \tag{5.133}$$

The flow is given by

$$\dot{p} = \sigma - \phi_{-1}(\sigma,t) = \sigma - (\underbrace{\phi_{-1}(s) + J^{-1}[\sigma - tF(s) - s]} + tF[\phi_{-1}(s)]) \tag{5.134}$$

$$= s - \phi_{-1} + (I - J^{-1})(\sigma - s)$$

Where underbraced term is the one second backwards time image of the point where (σ, t) first joins the sliding surface and the final term corrects for the fact that (σ, t)'s one second backwards time image is actually the $(1-t)$ second backwards time image

of this point. The transformed system flows according to the ODE $\dot{p} = \widehat{G}(p)$ whose linearization about $\Phi(s)$ is given by

$$\dot{p} = s - \phi_{-1}(s) + (J - I)\widetilde{J}^{-1}[p - \Phi(s)] \tag{5.135}$$
$$+\frac{\langle p - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle F(s) - F[\phi_{-1}(s)], (J\widetilde{J}^{-1})^T n\rangle}(J - I)\widetilde{J}^{-1}[F[\phi_{-1}(s)] - F(s)]$$

Expressions (119) and (124) agree on the switching surface $t = 0$ so the switch is differentiable.

Tail discontinuity

The flow is given by an ODE $\dot{x} = F(x)$ until reaching a set Σ where we instantaneously switch to an ODE $\dot{x} = H(x)$ as in (5.75). Solutions then flow for 1 second along the extended sliding surface Σ until reaching the vicinity of $\phi_1(s)$, where s is a grazing point. Again we use (σ, t) co-ordinates where $\sigma \in \Sigma$ is the current position of the point on the extended sliding surface and $t \ll 1$ is the branch index which tells us that its backwards time orbit stays on the extended sliding surface for $(1-t)$ seconds. If we attempt to expand Φ about $\phi_1(s)$ as in the basic slide tail-discontinuity we run into some difficulties, in particular since $H(s) = F(s)$ the second order terms cancel. We therefore include third order terms in t

$$\Phi(\sigma, t) = \Phi[\phi_1(s)] + \underbrace{\widetilde{J}J^{-1}[\sigma - \phi_1(s)]} + \int_0^t \phi_{-1+\tau}(x, t)d\tau - \int_0^t \phi_{-1+\tau}(x, 0)d\tau \tag{5.136}$$

Where the underbraced term is the derivative of Φ applied to σ restricted to the backwards flow branch where we stay on Σ for all time and the two integrals correct for the t seconds that we are not on the extended sliding surface.

In order to expand the first integral to third order in t we approximate the flow $\dot{x} = F(x)$ in the vicinity of s by

$$\dot{x} = F(s) + \underbrace{\frac{\delta F}{\delta x}|_s}(x - s) \tag{5.137}$$

Where we will denote the underbraced matrix A. This equations solution to second order in t is given by

$$x(t) = x(0) + t[A(x(0) - s) + F(s)] + \frac{t^2}{2}[A^2(x(0) - s) + AF(s)] \tag{5.138}$$

and for $x(0) = \phi_{-1+t}(\sigma) = [s + J^{-1}(\sigma - \phi_1(s)) + tF(s)]$

$$\int_{-t}^{0} x(t)dt = t[s+J^{-1}(\sigma-\phi_1(s))+tF(s)]-\frac{t^2}{2}[A[J^{-1}(\sigma-\phi_1(s))+tF(s)]+F(s)] \quad (5.139)$$

$$+\frac{t^3}{6}[A^2[J^{-1}(\sigma - \phi_1(s)) + tF(s)] + AF(s)]$$

Likewise we approximate $\dot{x} = H(x)$ by

$$\dot{x} = H(s) + \underbrace{\frac{\delta H}{\delta x}|_s}(x - s) \quad (5.140)$$

for the second integral. To obtain

$$\Phi(\sigma, t) = \Phi(\phi_1(s)) + \widetilde{J}J^{-1}[\sigma - \phi_1(s)] + \frac{t^3}{3}(\frac{\delta H}{\delta x}|_s - \frac{\delta F}{\delta x}|_s)F(s) \quad (5.141)$$

which is invertible in the usual way

$$\frac{t^3}{3} = \frac{\langle \Phi(\sigma, t) - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle (\frac{\delta H}{\delta x}|_s - \frac{\delta F}{\delta x}|_s)F(s), (J\widetilde{J}^{-1})^T n\rangle} \quad (5.142)$$

where n is normal to Σ at $\phi_1(s)$.

$$\sigma = \phi_1(s) + J\widetilde{J}^{-1}[\Phi(\sigma, t) - \Phi[\phi_1(s)] - \frac{t^3}{3}(\frac{\delta H}{\delta x}|_s - \frac{\delta F}{\delta x}|_s)F(s)] \quad (5.143)$$

The flow is given by

$$\dot{p} = \sigma - \phi_{-1}(\sigma, t) = \sigma - [\underbrace{s + J^{-1}[\sigma - \phi_1(s)]} + \frac{t^2}{2}(\frac{\delta F}{\delta x}|_s - \frac{\delta H}{\delta x}|_s)F(s)] \quad (5.144)$$

where the underbraced term is the one second backwards time image of x restricted to the branch where we stay on the sliding surface for all time and the final term corrects for the t seconds of flow off the sliding surface. So that the transomed system flows according to the ODE $\dot{p} = \widehat{H}(p)$ whose expansion about $\Phi[\phi(s)]$ is given by

$$\dot{p} = a + B(p - \Phi[\phi_1(s)]) + c\sqrt[\frac{3}{2}]{\langle p - \Phi[\phi_1(s)], d\rangle} \quad (5.145)$$

where

$$a + B(p - \Phi[\phi_1(s)]) = \phi_1(s) - s + (J - I)\widetilde{J}(p - \Phi[\phi_1(s)]) \quad (5.146)$$

$$-\frac{\langle \Phi(\sigma, t) - \Phi(s), (J\widetilde{J}^{-1})^T n\rangle}{\langle (\frac{\delta H}{\delta x}|_s - \frac{\delta F}{\delta x}|_s)F(s), (J\widetilde{J}^{-1})^T n\rangle}(J - I)\widetilde{J}(\frac{\delta H}{\delta x}|_s - \frac{\delta F}{\delta x}|_s)F(s)$$

and

$$c\sqrt[\frac{3}{2}]{\langle p-\Phi[\phi_1(s)],d\rangle}=\frac{1}{2}(\frac{\delta F}{\delta x}|_s-\frac{\delta H}{\delta x}|_s)F(s)\sqrt[\frac{3}{2}]{3\frac{\langle\Phi(\sigma,t)-\Phi(s),(J\widetilde{J}^{-1})^Tn\rangle}{\langle(\frac{\delta H}{\delta x}|_s-\frac{\delta F}{\delta x}|_s)F(s),(J\widetilde{J}^{-1})^Tn\rangle}} \tag{5.147}$$

So that as in the basic sliding case orbits converge differentially and in finite time to the image of the extended sliding surface $\Phi(\Sigma,0)$ except that in this grazing case the distance decays like $(t-c)^3$ due to the power of $\frac{2}{3}$ in the flow.

5.4 Computer example

The Moving Average Filter can easily be applied to time series data from a computer simulation of a non-smooth system. Since this data is equivalent to unsmoothed output from the smoothed system we can use it to study the systems obtained by applying the Moving Average Transformation to non-smooth systems.

5.4.1 Friction oscillator

Systems with static and dynamic friction can exhibit sliding behavior, there is a (so far as possible) humorous mismatch in terminology here since mathematical sliding in the sense of section 3.3 actually corresponds to an object being stuck to a surface by static friction whilst sliding along experiencing dynamic friction has no mathematical sliding associated with it at all! In our model we consider a mass on a linear spring subjected to a periodic force resting on a rough moving belt with a piecewise linear friction force

$$Fric(y)=\begin{cases} ay+b & y>0 \\ -ay-b & y<0 \end{cases} \tag{5.148}$$

where y is the velocity of the mass relative to the surface of the belt. The force acting on the block is given by

$$F(x,\dot{x},t)=kx-d\dot{x}+l\sin(\omega t)+Fric(\dot{x}-u) \tag{5.149}$$

The state space of this dynamical system is three dimensional (position,velocity,forcing-phase). We will embed the phase variable as an interval $[0,2\pi]$ rather than on the

circle as it enables us to embed the the whole state space in $\mathbb{R}^3$ and the jump from 2π to 0 gives us some discontinuities in the flow which we can resolve with the transformation.

Procedure

For a given set of parameter values we simulate the oscillator on a computer and record a time series in the three dimensional state variable. The recording is just a long sequence $(x_i, \dot{x}_i, \tau_i)_{i=1}^N$ where we use a time-step of $0.05s$. The smoothing filter is just a summed average applied to the now discrete time data.

$$\Phi(x, \dot{x}, \tau)_i = \frac{1}{40} \sum_{j=i-39}^{i} (x_j, \dot{x}_j, \tau_j) \tag{5.150}$$

So that we are averaging over a period of $2s$. The accuracy of our routine is actually slightly better than this as we can approximate the average between time steps during the integration routine which improves the resolution of the sum. To obtain a differentiable time series we simply apply the same filter a second time to the data.

Data

We use the following parameters $k = 1.2$, $d = 0.0$, $w = 1.02$, $l = 1.9$, $u = 3.4$, $a = 0.04$ and $b = 0.1$ whose dynamics exhibit a grazing orbit that gives rise to chaotic dynamics on what appears to be a fully 2 dimensional attractor. We plot the time series for the three systems, discontinuous, continuous/non-differentiable and differentiable, see Fig 8,1.

Results

Both applications of the filter give us useful insight into the original systems dynamics. When we transform the system from a discontinuous system to a continuous system we change the topology of the attractor. As outlined in the introduction we claim that the topology of the continuous system gives a better characterization of the dynamics and that is the case here. Homologically the discontinuous system is equivalent to

two points whereas the continuous system is a figure of eight, the two possible loops in the attractor provide the topological mechanism for chaos, see Fig 8,b.

On the second application of the filter we arrive at a differentiable system which can be thought of as an ODE with continuous RHS. We can therefore locally approximate this ODE from our data and use this to compute lyupanov exponents along the flow. The clear positive exponent shows that we have a chaotic system, see Fig 8,c.

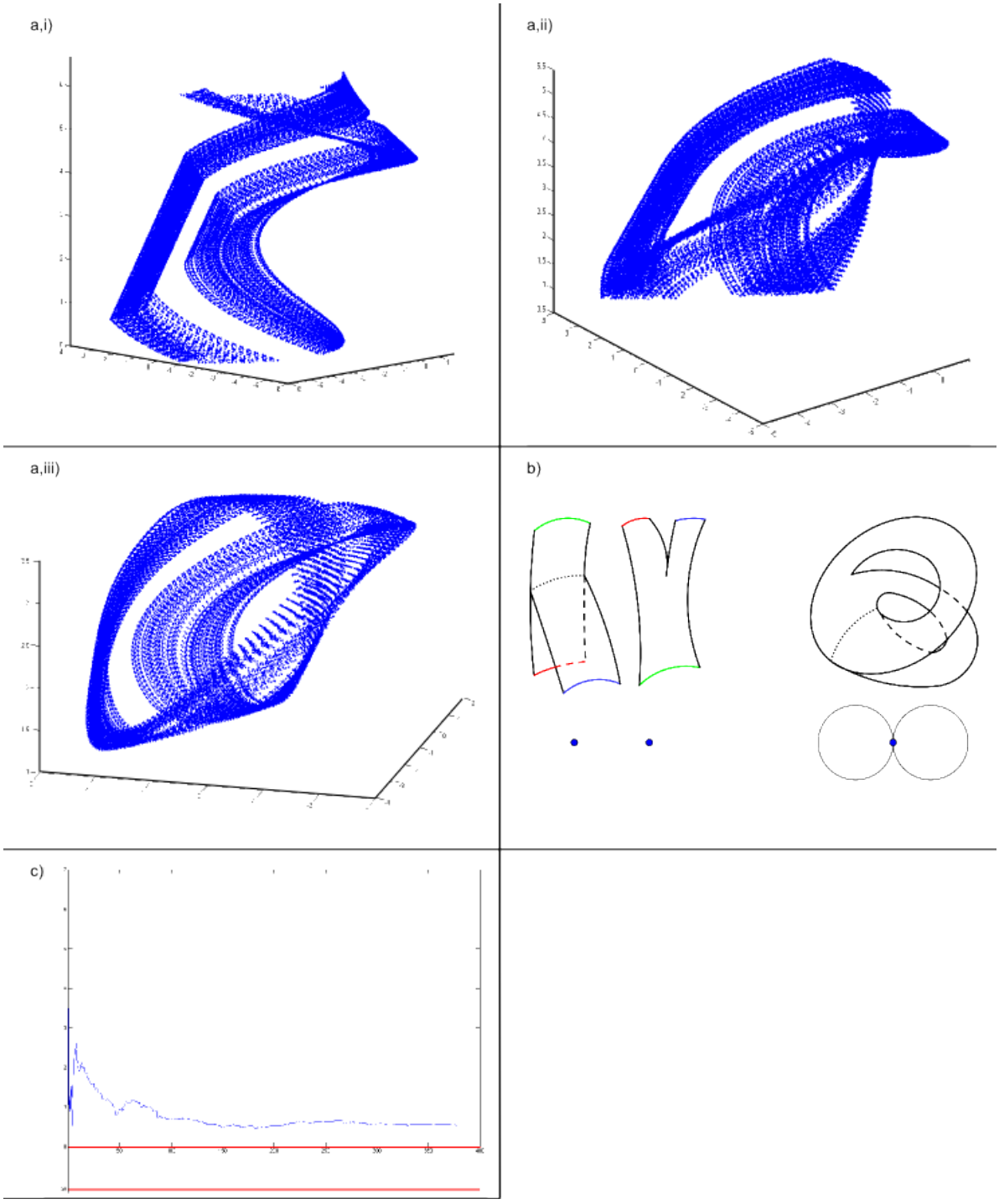


Figure 5.8: a) Time series recorded from i) Original discontinuous system, ii) Once transformed continuous system, iii) Twice transformed differentiable system, b) Convergence of Lyupanov exponent calculation, c) Topology of invariant set for discontinuous and transformed continuous system.

Local behavior of transformed systems

The original discontinuous system contains three jumps, three slides (green) and one sliding graze. Through the averaging these discontinuities are transformed into smoother discontinuities as in section 3. We can easily identify the discontinuities in the data during the simulation stage.

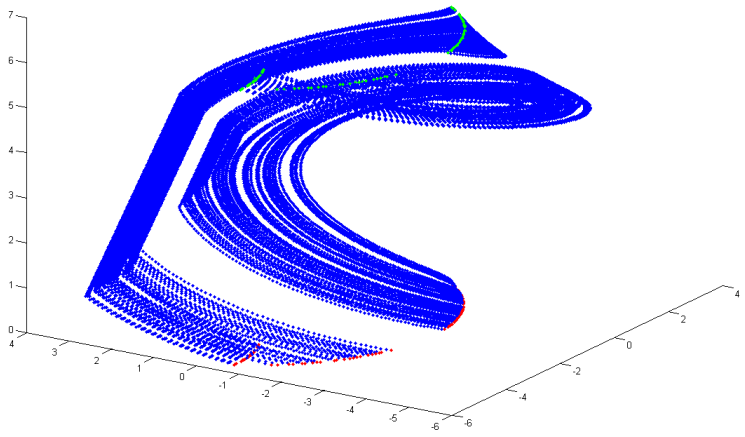


Figure 5.9: Original system with discontinuous jumps are highlighted. Data just before the discontinuity in Green and after in Red

As in section 5.3 the discontinuous jumps in the original system are transformed into pairs of non-differentiable switches, a second application of the transformation smoothes these into differentiable switches. See figs 9,10,11.

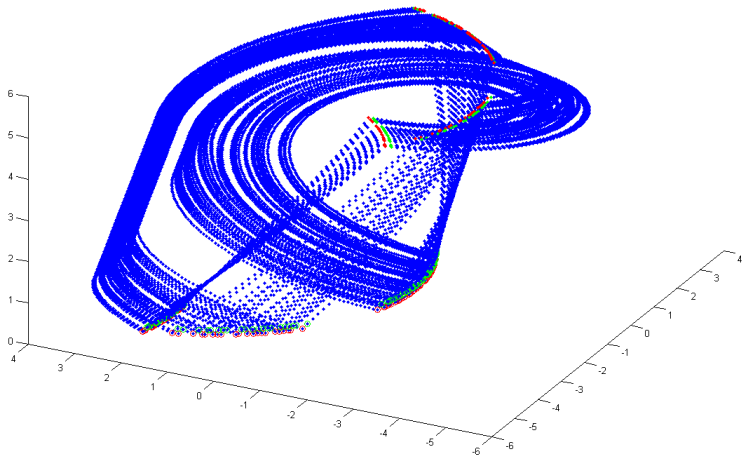


Figure 5.10: Once smoothed system with image of discontinuous jumps highlighted. Data just before the discontinuity in Green and after in Red, Head with dots and tail with circles.

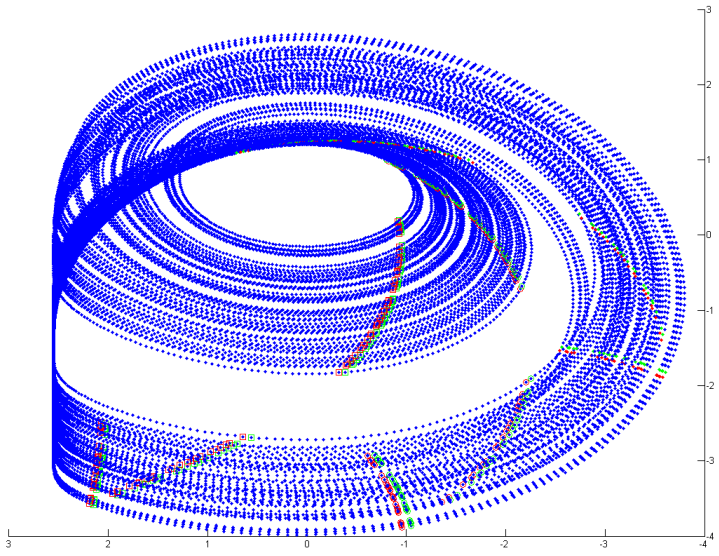


Figure 5.11: Twice smoothed system with image of discontinuous jumps highlighted. Data just before the discontinuity in Green and after in Red, head with dots, mid with circles and tail with squares.

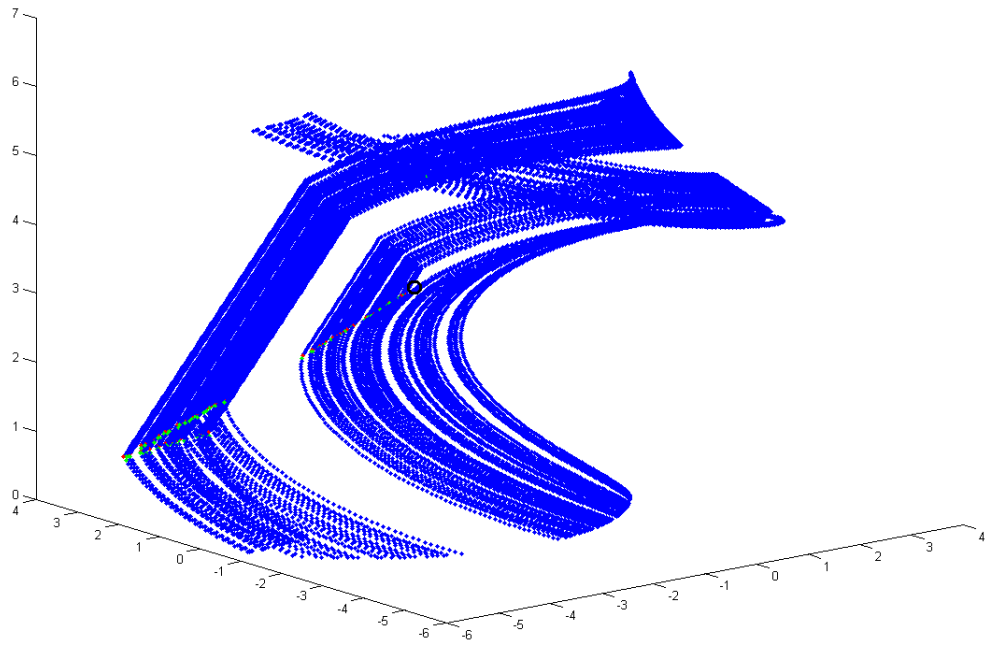


Figure 5.12: Original system with non-differentiable slide highlighted. Data just before the discontinuity in Green and after in Red

As in section 5.3 the non-differentiable slides in the original system are transformed into pairs of non-smooth discontinuities each consisting of a differentiable switch followed by a differentiable slide where square root terms in the flow give rise to parabolic orbits which flow into the transformed sliding surface tangentially. See Figs 12,13,14.

This system contains a sliding graze discontinuity which acts as a separatrix between the two different loops observed in the orbits. The grazing point itself does not appear to be in the closure of the attractor.

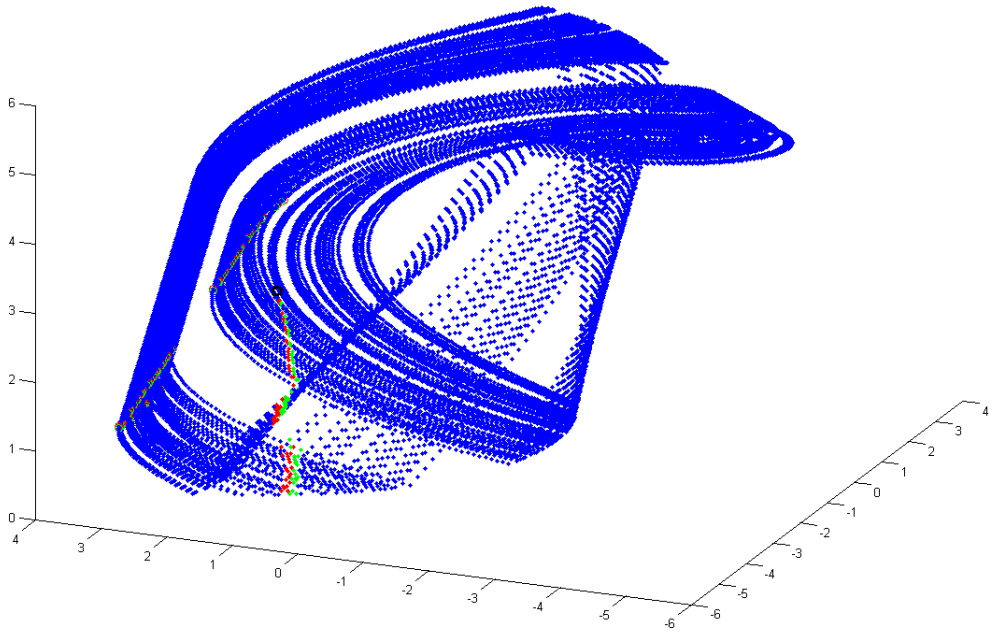


Figure 5.13: Once smoothed system with image of non-differentiable slide highlighted. Data just before the discontinuity in Green and after in Red head with dots, mid with circles and tail with squares.

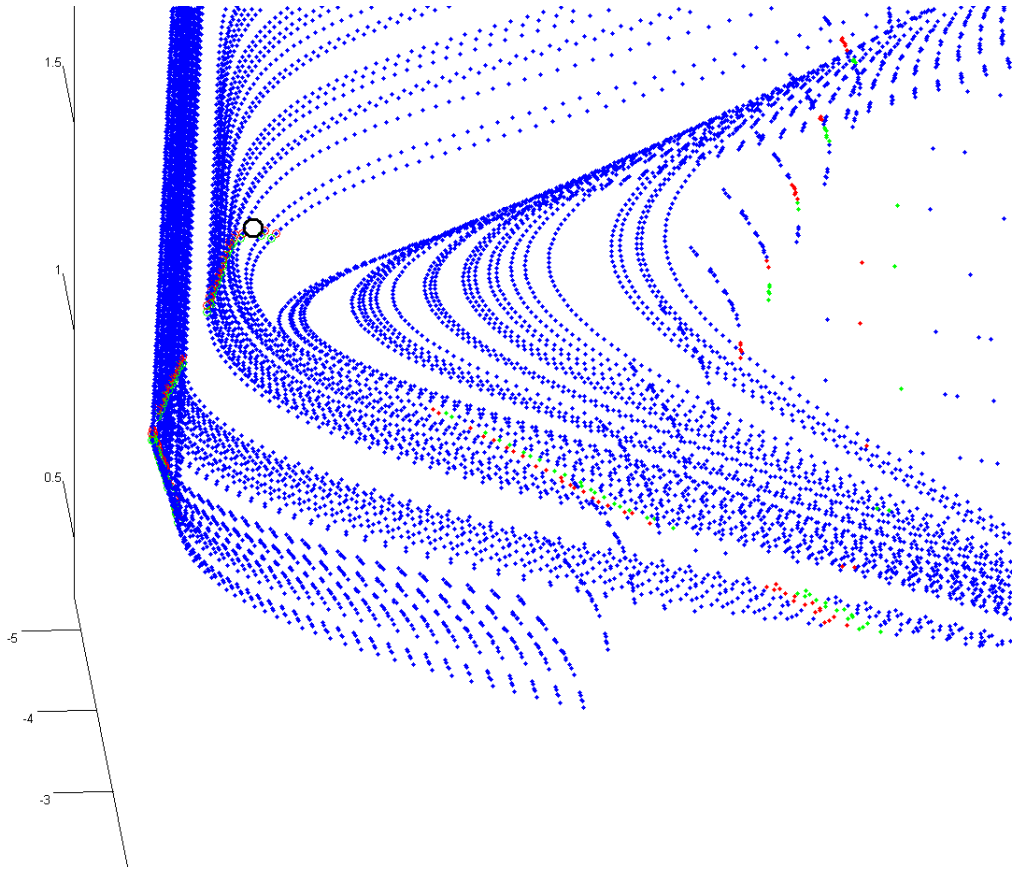


Figure 5.14: Detail of Tail discontinuity for smoothed slide.

Remarks on averaging period

In the theory laid out earlier in the chapter the length of the averaging period really didn't matter. If we set

$$\Phi_\alpha(x) = \int_{-\alpha}^{0} \phi_\tau(x) d\tau \tag{5.151}$$

then the Φ_α transformed systems will all be topologically equivalent and admit the same sort of local descriptions at the images and tail images of discontinuities. In the practical world of numerics this is obviously not the case. There are two strategic factors in play. Firstly if we make α to large then we loose a lot of information when we take the average as Φ_α will be too contracting, mix this with some natural round up error and the transformed system will begin to look like a smooth blob with noise and little resemblance to the original system. On the other hand if α is too small then smoothed discontinuities will still have very large changes in second derivatives

which will lead to more numerical errors.

A theoretical basis for choosing α is still a work in progress but for non-smooth oscillators like this one an averaging period roughly half of a typical orbit period seems to work well as a rule of thumb. Moving averages of Markov chains with exponentially decaying memory of the form

$$\Phi(x) = \int_{-\infty}^{0} \phi_\tau(x) e^{\beta\tau} d\tau \tag{5.152}$$

are of some interest in approximating the behavior of coupled systems with switching network architecture. Here the entropy of the Markov chain provides a natural time scale when considering the choice of β [12], perhaps something like that can be said here.

Remarks on alternative methods

Of course the results outlined above could be obtained without making use of the Moving Average Transformation, traditional methods are however quite difficult to apply in a systematic way. In order to identify the topology of the continuous system it would be necessary to identify points before and after the time reset jumps as being connected then add these connections to something like the rips complex of the series.

To compute the lyupanov exponent from the data as the integral of the derivative along the flow it would be necessary to divide the points in the series up into 5 categories

1. Points that are not sliding and not about to start siding or time reset.
2. Points that are not sliding and are about to start sliding
3. Points that are not sliding and are about to time reset
4. Points that are sliding and are not about to time reset
5. Points that are sliding that are about to time reset

The vector field/saltation matrix can then be adequately approximated at a point in category i say, by a local approximation using other close by points in that same

category. This wouldn't be too difficult from a computer simulation like this where it would be easy to categorize points as such but would be extremely difficult when dealing with experimental data where it would be impossible to verify the categorization. Of course it is possible to compute lyupanov exponents from data by constructing a return map which placed away from the discontinuity would not be too problematic, however this is very far from systematic doesn't give us an approximation of the vector field.

5.5 Conclusion

We have shown that the Moving Average Transformation can be used to transform discontinuous or non-differentiable systems into dynamically equivalent differentiable systems. That stability and orbit structure are dynamical equivalence invariants follows from the fact that away from the discontinuities we have a standard smooth change of variables and on the discontinues the theory in section 3 shows how in a quite natural way this information is integrated into the smoothed system. We have also demonstrated that the transformation can be applied numerically especially well to time-series data where we can replace the state space transformation with a simple finite impulse response (FIR) filter on the data. There are several questions left open all of which I hope to address in future work

5.5.1 Numerical implementation

There are many questions here. Firstly as discussed in 4.1.4 how should we choose the length of the averaging period? Whilst the time-series based method is very easy to implement it is not necessarily the best approach, for example we are only able to collect reasonably dense data from attracting objects. How should one best go about numerically implementing a method like that used in section 2? Then there are all the usual questions associated with numerics such as cost, the effect of noise (FIR filters like the moving average filter is routinely used to remove high frequency noise in signal processing) and rounding error which could be very important here

especially if we want to use a longer averaging period where the more contracting transformation can loose sensitivity.

In section 4 we where able to compute the Lyupanov spectrum of the Friction Oscillator from a time-series using our new technique. Do the benefits of this approach which include systematic implementation, use of the whole vector field and implementation of well tested/reasearched standard smooth methods outweigh the potential problems such as the introduction of systematic numerical error through the transformation and increased dimension through use of delay maps?

5.5.2 More applications

This chapter has focused on computing the stability of non-smooth systems by smoothing them then applying standard techniques that really on differentiability which would not have been applicable before the smoothing. Of course stability calculations are not the only standard techniques that rely on smoothness. What other standard techniques can be applied to the smoothed systems to get results that would have been impossible or more difficult to obtain from the original non-smooth system? Numerical continuation for non-smooth systems is very difficult because of the non-differentiability [4]. Can we use our smoothing technique to do this in a simpler more systematic way?

5.5.3 Generic Smoothed non-smooth systems

Non-smooth systems can generically exhibit bifurcation structures which would be impossible or of high codimension in the space of smooth systems. Since dynamically equivalent systems have the same dynamics and therefore bifurcate in the same way the Moving Average Transformation provides us with a family of smooth systems (the smoothed non-smooth systems) which behave in an atypical way. How can we characterize these systems? What does a generic smoothed system look like and what insight into non-smooth dynamics does this answer give us, if any? we conjecture that a generic smoothed system can be expressed as a generic smooth flow

on special type of branching manifold that splits up and folds back into itself along non-transverse intersections with grazes and slides respectively responsible for the two types of irregularity.

Bibliography

[1] Alart P, Maisonneuve O, Rockafellar R.T. (2006). Nonsmooth Mechanics and Analysis. *Springer*

[2] Bernardo M, Budd C, Champneys A.R, Kowalczyk P. (2008). Piecewise-smooth Dynamical Systems. *Springer.*

[3] Kunze M, K*ü*pper, T. (2001). Nonsmooth dynamical systems: An overview. *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems.* pages 431-452.

[4] Arango I, Taborda J. A. (2007). Continuation of nonsmooth bifurcations in filippov systems using singular point tracking. *International Journal of Applied Mathematics and Informatics.* Issue 1, volume 1, pages 36-49.

[5] Llibre J, Da Silva P.R, Teixeira M.A. (2007). Regularization of Discontinuous Vector Fields on $\mathbb{R}^3$ via Singular Perturbation. *Journal of Dynamics and Differential Equations.* Volume 19, number 2, pages 309-331.

[6] Gang T, Frank L. (2001). Adaptive Control of Nonsmooth Dynamic Systems. *Springer.*

[7] Di Bernardo M, Hogan S.J. (2010) Discontinuity-induced bifurcations of piecewise smooth dynamical systems. *Phil. Trans. R. Soc. A.* Volume 368, number 1930, pages 4915-4935.

[8] Kunze M, K*ü*pper T. (1996). Qualitative bifurcation analysis of a non-smooth friction-oscillator model. *ZAMP.* Volume 48, pages 87-101.

[9] Huke J. (1993) Embedding nonlinear dynamical systems: A guide to Takens' theorem. *Technical Report DRA, Malvern.*

[10] Filippov A.F. (1960). Differential equations with discontinuous right-hand side. *Mat. Sb. (N.S.)* Volume 51, number 1, Pages 99-128.

[11] Darbyshire G. and Broomhead D.S. (1996). The calculation of Liapunov exponents from time series data. *Physica D.* Volume 89, pages 287-305.

[12] Skufca J.D, Bollt E.M. (2004). Communication and Synchronization in Disconnected Networks with Dynamic Topology Moving Neighborhood Networks. *Mathematical Biosciences and Engineering.* Volume 1.

Chapter 6

Future plans

Thanks to an EPSRC doctoral prize award I will be able to continue my work in Manchester for the 2011/2012 academic year. Principally developing applications of my smoothing technique I hope to continue much of the work presented in my thesis. I am particularly keen to collaborate with engineers working on non-smooth systems as my smoothing method's numerical stability and ability to deal with data will hopefully make it useful in 'real life' applications. I also look foreword to investigating more realistic models of parallel computation and continue to take a keen interest in max-plus algebra through the departments tropical mathematics group.

Finally I would like to thank again my wonderful advisors; David Broomhead, Paul Glendining and Mark Muldoon. I have thoroughly enjoyed my time with them as a PhD student at Manchester and look forward to continuing to work with them this coming year.

Thank you for reading!