Mathematical Epidemiology

Gauthier Sallet

Mathematical Epidemiology

Gauthier Sallet
Emeritus Professor
Institut Élie Cartan, UMR CNRS 7502
Université de Lorraine

March 2018

Contents

0.1 Preface

The origin of these lectures notes comes from a series of courses taught in the University of Pretoria in Match 2018. There are, now, many books in mathematical epidemiology [18, 17, 25, 10, 64, 92] to cite a few. These lectures notes consider only finite dimensional deterministic system, ODE for short.

In these lectures notes we have addressed some issues which are not commonly treated elsewhere.

1. We have given detailed exposition on Lyapunov and LaSalle techniques with examples from the literature.

   The reduction techniques of Vidyasagar are given. They are used throughout these lectures, simplifying the models by dimension reduction.

2. The exposition on $\mathcal{R}_0$ is now classical, but we try to clear the notion of threshold in relation with Varga's theorem;

3. We think that our exposition on Monotone systems with applications to large scale epidemiological systems is original. All the results are proven in details, results which are dispersed in the literature.

4. The linear chain trick is exposed and show how to deal with delays only with ODE

5. The problem of parameters, identification and identifiability is introduced with a control theory approach.

I would like to thanks the department of mathematics of UP, the staff and their students for their warm welcome and to permit this opportunity to give these lectures.

Chapter 1

Introduction and Important Concepts

1.1 Mathematical modeling of infectious diseases

Mathematical epidemiology has a long history. In 1760 Daniel Bernoulli presents a study on smallpox (called "petite vérole") under the title "Essai d'une nouvelle analyse de la mortalite caus?e par la petite v?erole & des avantages de l'inoculation pour la pré?venir." These study is often quoted as the first epidemiological model. In introduction of these study Bernoulli says

> je souhaite seulement dans une question qui regarde de si pr?s le bien de l'humanit?, on ne d?cide rien qu'avec toute la connaissance de cause qu'un peu d'analyse & de calcul peut fournir

(*I only wish in a matter that looks so closely at the good of humanity, we decide nothing with all the knowledge that a little analysis & calculation can provide.*)
Actually the foundations of mathematical epidemiology were laid not by mathematicians but by physicians in the beginning if the twentieth century.

P. D.En'ko's paper (1889) "on the course of epidemics of some infectious diseases, VRAC 1889" was the first to discuss the elements of a genuine epidemic model, namely the chain binomial. The foundations of the entire approach to epidemiology based on compartmental models were laid by public health physicians such as Sir Ross [78, 79], R.A., W.H. Hamer, A.G. McKendrick and W.O. Kermack [57] between 1900 and 1935.
A particularly instructive example is the work of Ross on malaria. Sir Ronald Ross was awarded the second Nobel Prize in Medicine in 1902 for his demonstration of the dynamics of the transmission of malaria between mosquitoes and humans. Ross waged a constant and often acrimonious battle for the acceptance of what he called his "mosquito theorem" p22 of [77] :

> The word theorem is used here in its correct sense as expressing not a mere speculation, but a body of established fact.

The implication of this theorem is that a reduction of *Anopheles* population is a mean to prevent Malaria. The argument against Ross's ideas were

- it is impossible to totally eradicate the mosquitoes in an area (Ross admitted this);
- thus there will always be some mosquitoes remaining (Ross admitted this);
- thus malaria transmission will continue, and vector control is a waste of time and effort (here Ross disagreed)

The fallacy of such an argument is now well established : you must think quantitatively ! To quote Ross again [78] :

> The mathematical treatment adopted in section 28 has been met with some questioning by critics. Some have approved of it, but others think that it is scarcely feasible owing to the large numbers of variables which must be considered. As a matter of fact all epidemiology concerned as it is with the variation of disease from time to time or from place to place, must be considered mathematically, however many variables are implicated, if it is to be considered scientifically at all. **To say that a disease depends upon certain factors is not to say much**, until we can also form an estimate as to how largely each factor influences the whole result. And the mathematical method of treatment is really nothing but the application of careful reasoning to the problems at issue

Ross insists on the qualitative nature of the studies.

It was the challenge of convincing the world that mosquito control was a practical public health undertaking that stimulated Ronald Ross to develop his model in 1911.

The mechanism of transmission of infections is now known for most diseases. Generally, diseases transmitted by viral agents, such as influenza, measles, rubella (German measles) and chicken pox, confer immunity against reinfection, while diseases transmitted by bacteria, such as tuberculosis, meningitis and gonorrhea, confer no immunity against reinfection. Other diseases, such as malaria,are transmitted not directly from human to human but by vectors, agents (usually insects) that are infected by humans and subsequently transmit the disease back to humans.

Mathematical epidemiology differs from most sciences as it does not lend itself to experimental validation of models. Experiments are usually impossible and would probably be unethical. This gives great importance to mathematical

models as a possible tool for the comparison of strategies to plan for an anticipated epidemic or pandemic, and to deal with a disease outbreak in real time [18].

Are mathematics useful in epidemiology ? we will quote Ross again. Ronald Ross is not the first to use mathematics to study some problems in epidemiology. But it can be said that Ross is the first to systematically use the mathematical approach . In the introduction of *An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I* Ross writes

> The whole subject (i.e., Epidemiology) is capable of study by two distinct methods which are used in other branches of science, which are complementary of each other, and which should converge towards the same results – the *a posteriori* and the *a priori* methods. In the former we commence with observed statistics, endeavour to fit analytical laws to them, and so work backwards to the underlying cause (as done in much statistical work of the day) ; and in the latter we assume a knowledge of the cause, we construct our differential equations on that supposition, follow up the logical consequences, and finally test the calculated results by comparing them with the observed statistics

Ross is the first to use the *a priori* method and the cited text, dating back to 1916 ! It is a good introduction to a course on modelling.

1.2 Deterministic epidemic models : compartmental approach

Dynamic models of many processes in pharmacokinetics;metabolism, epidemiology, ecology, and other areas are derived from mass balance considerations

As a result, these models lead to particular systems of ordinary differential equations–many of them nonlinear–that are called compartmental systems.

The equations of compartmental systems are subject to such strong structural constraints that it seems likely that their solutions may also be strongly constrained.

The conservation law that dominates such systems is the law of conservation of mass : there are also called mass balance systems. A compartment is an amount of some material that is kinetically homogeneous.

By kinetically homogeneous we mean the material of a compartment is at all times homogeneous; any material entering it is instantaneously mixed with the material of the compartment.

- A space or a region limited by barriers
- Or a substance or a physical quantity ,without precise localization

An example is the presence of lead in an living organism :

Lead in an organisme gives rise to a disease : lead poisoning (also call saturnism)

a part of lead absorbed is excreted, but the remaining accumulate essentially in bones; 80 to 95%. In bones lead has a mean half life of 20–25 years.

Lead also accumulate in liver, kidneys, brain with irreversible and severe effects in the organism. Modeling lead poisoning leads to consider 3

compartments : blood in which lead is transported, soft tissues and finally bones.

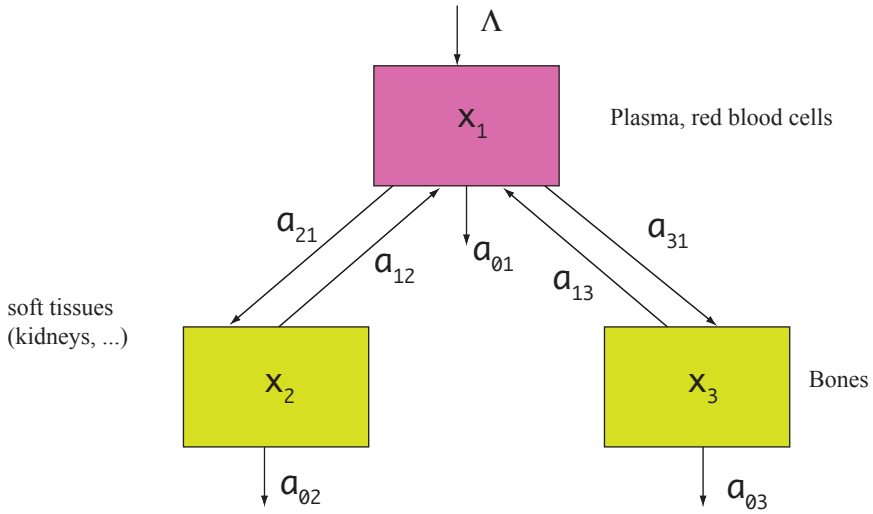


Figure 1.1: Flow graph of lead poisoning

1.2.1 Compartmental equations

The box in the following figure represents the i-th compartment of an n compartment system.

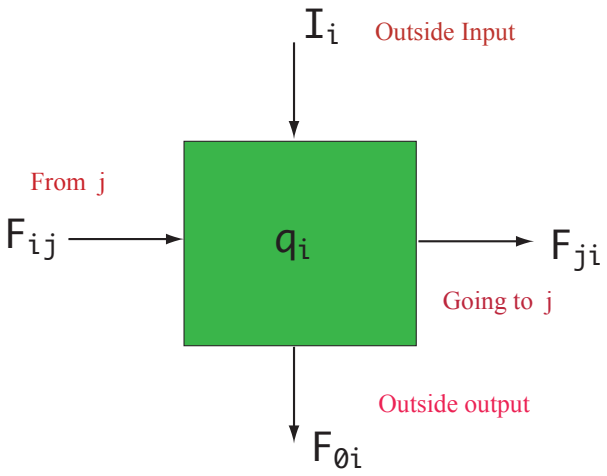


Figure 1.2: A comparment

Arrows represents the input and output flows in the compartment.

$$\dot{q}_i = I_i - F_{0i} + \sum_{j \neq i} F_{ij} - F_{ji}$$

Inputs : the flows into the compartment

$$I_i(t)\,\Delta t + \left(\sum_j F_{ij}\right)\Delta t$$

Outputs : the outflows leaving the compartment

$$F_{0i}\,\Delta t + \left(\sum_{j \neq i} F_{ji}\right)\,\Delta t$$

Instantaneous mass balance equations

$$\dot{q}_i(t) = I_i(t) \; + \left(\sum_{j \neq i} F_{ij}\right) - F_{0i} \; - \left(\sum_{j \neq i} F_{ji}\right)$$

The functions F_{ij} and I_i are flows : quantity of material by unit of time.

The functions I_i, F_{0i}, F_{ij} can be functions of $q_1, \ldots, q_n$ and possibly t.
So we can write the functions $F_{ij}(t, q)$
These are nonnegative quantity $I_i \geq 0$, $F_{0i} \geq 0$, $F_{ij} \geq 0$
If there is no material un a compartment, nothing can leave the compartment.
Mathematically

$$q_i = 0 \Rightarrow F_{ji} = 0 \text{ and } F_{0i} = 0$$

To summarize we have

- $$F_{ij} \geq 0 \;\; F_{0i} \geq 0 \;\; I_i \geq 0$$

- $$q_i = 0 \Rightarrow F_{ji} = 0 \text{ et } F_{0i} = 0$$

Now we will suppose that these functions are $\mathcal{C}^1$

Proposition 1.2.1 *If f is a function from $\mathbb{R}^n$ into $\mathbb{R}^m$ of class $\mathcal{C}^k$, s.t. $f(x^*) = 0$ there exists $A(x)$ of class $\mathcal{C}^{k-1}$, from $\mathbb{R}^n$ in matrices $m \times n$ such that for any $x \in \mathbb{R}^n$ we have*

$$f(x) = A(x)\,(x - x^*)$$

Proof
we consider $\mathcal{C}^1$ from $\mathbb{R}$ into $\mathbb{R}^m$

$$\varphi(t) = f(x^* + t\,(x - x^*))$$

$$f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(s)\, ds$$
$$= \int_0^1 Df(x^* + s\,(x - x^*)\,.\,(x - x^*)\; ds$$
$$= \left(\int_0^1 Df(x^* + s\,(x - x^*)\, ds \right)\; (x - x^*)$$

Then $A(x) = \left(\int_0^1 Df(x^* + s\,(x - x^*)\, ds \right)$ of class $\mathcal{C}^1$.

■

Consequence for our functions there exists a function f_{ij} such that

$$F_{ij} = f_{ij}\, q_j$$
$$\dot{q}_i = I_i - F_{0i} + \sum_{j \neq i} F_{ij} - F_{ij}$$
$$\dot{q}_i = -\left(f_{0i} + \sum_{j \neq i} f_{ji} \right) q_i + \sum_{j \neq i} f_{ij}\, q_j + I_i$$

We now introduce a matrix A defined by

- $A(i,j) = f_{ij}$
- $A(i,i) = -f_{0i} - \sum_{j \neq i} f_{ji}$
- and the vector $\mathbf{I} = (I_1, \cdots, I_n)^T$

The equations can now be written in a linearized way

$$\dot{\mathbf{q}} = A\,\mathbf{q} + \mathbf{I}$$

Functions f_{ij} are called *fractional transfer coefficients*. Dimension of these functions are t^{-1}. Depending generally for q and t.

The entries of the matrix A have some properties : the off-diagonal entries are nonnegative.

Definition 1.2.1 (Metzler Matrix) *A matrix A whose off diagonal entries are nonnegative, i.e. if $i \neq j$ then $a_{ij} \geq 0$ is called a Metzler matrix.*

But we have more properties for our matrices A

$$\dot{\mathbf{q}} = A(t, \mathbf{q})\,\mathbf{q} + \mathbf{I}(\mathbf{t}, \mathbf{q})$$

a diagonal entry is given by $A(i,i) = -f_{0i} - \sum_{j \neq i} f_{ji} \leq 0$. In other words $A(i,i)$ is equal to minus the sum of entries of columnn i and subtracting the term $-f_{0i}$. Then for the matrix A, each column sum is non positive.
A Metzler matrix, which satisfies that the column sum are nonpositive (this implies that the diagonal terms are non positive) diagonale) is called a compartmental.

1.2.2 Graphic representations

A standard graphical representation of compartmental systems uses nodes for compartments and directed arcs labeled with fractional transfer coefficients for transfers between compartments, and for excretions.

Inputs are labeled with the input function. Such representations are called compartmental system connection diagrams or simply connection diagrams or flow-graphs. Actually these are digraphs with weight on each arc : these are also called un Coates graph.

1.2.3 An example : The Kermack?McKendrick Model

We formulate our descriptions as **compartmental models**, with the population under study being divided into compartments and with assumptions

about the nature and time rate of transfer from one compartment to another.

Diseases that confer immunity have a different compartmental structure from diseases without immunity

In order to model such an epidemic we divide the population being studied into three classes labeled S, I, and R. We let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t. $I(t)$ denotes the number of infected individuals, assumed infectious and able to spread the disease by contact with susceptibles. $R(t)$ denotes the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection. Removal is carried out either through isolation from the rest of the population or through immunization against infection or through recovery from the disease with full immunity against reinfection or through death caused by the disease.In formulating models in terms of the derivatives of the sizes of each compartment we are assuming that the number of members in a compartment is a differentiable function of time. This may be a reasonable approximation if there are many members in a compartment, but it is certainly suspect otherwise. The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933

$$\begin{cases} \dot{S} &= -\beta\, S\, I \\ \dot{I} &= \beta\, S\, I - \gamma\, I \\ \dot{R} &= \gamma\, I \end{cases} \tag{1.1}$$

Actually, in the paper of Kermack and McKendrick, S, I and R are area densities.

We have the following hypothesis

- Constant population.

- All individuals are equally susceptible
- Infection leads to death or complete recovery with permanent immunity

The flow graph

Figure 1.3: Flow graph of Kermack-McKendrick model

Explanation : the production of new infective is proportional to the product of densities S and I. Intuitively $S\,I$ is the probability of encounter.

This is the mass action law, when the variables are densities.

Another way to formulate is : we assume that any individual makes β adequate contact by unit of time with others. If N is the total population and S, I and R prevalence (i.e., % relatively to the population), we have $I\,N$ infectious individuals, then the number of encounter of an infective by unit of time is $(\beta\,N)\,I$. Among this encounters, only the encounters with a susceptible, will produce a new infective

Then the number of new infections in unit time per infective is equal to $(\beta\,N)\,I\,S = \beta, N\,S\,I = \tilde{\beta}\,S\,I$

Note : the hypothesis of homogeneity.

1.2.4 Transfer rates

For the simple Kermack-McKendrick model described in the previous section, we assumed that the recovery rate, or the rate of transfer from compartment I to R, is given by γI . This is equivalent to assuming the following:

(H) the fraction of the infectious population that recovers per unit time is a constant.

Proportional transfer rates as assumed in (H) are often used for transfers between compartments in simple compartmental models. However, we need to understand that this is only one of many assumptions we can make about population transfers.

In fact, our assumption that recovery rate is in proportion to the size of the infectious population is by no means universal. In the following, we develop a better mathematical understanding of the proportional transfer rate, and consider other possible alternatives. Consider a general compartment C of total population size $N(t)$, where individuals leave the compartment at a rate $rN(t)$ $(r > 0)$. Then the size $N(t)$ satisfies

$$\frac{dN(t)}{dt} = -r\,N(t), \quad r > 0,$$

and thus $N(t) = N_0\,e^{-r\,t}$, or

$$\frac{N(t)}{N_0} = e^{-r\,t}.$$

Therefore, $e^{-r\,t}$ gives the fraction of the population that remains in the compartment C. In probability terms, $e^{-r\,t}$ is the probability of an individual entering C at time $t = 0$ and remaining in C at time $t > 0$. Since we are interested in the population transfer out of C, we consider

$$F(t) = \begin{cases} 1 - e^{-r\,t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

which gives the fraction of the population that has left C during the time period $[0, t)$, or the probability of an individual who has left C during $[0, t)$. Here we see that $F(t)$ has the characteristics of a probability distribution.

In fact, let X denote the random variable of the residence time of an individual in compartment C, the time period from entrance to exit, we see that

$$F(t) = \text{Prob}[\, X \leq t \,].$$

In other words, $F(t)$ is the probability distribution function of individual residence time in C, and it satisfies the following properties:

- $F(t) \geq 0$
- $F(t) \to 0$, as $t \to -\infty$
- $F(t) \to 1$, as $t \to +\infty$

Now we see that the assumption of proportional exit rate is the same as the following: (H0) the residence time of an individual in compartment C has an exponential distribution. We can also describe the random variable X in terms of probability density function $f(t) = \dfrac{d}{dt} F(t)$, namely:

$$f(t) = \begin{cases} r\, e^{-r\,t}, & t \geq 0 \\ 0, & t < 0, \end{cases}$$

with the following properties

- $f(t) \geq 0$

- $\int_{-\infty}^{+\infty} f(t)\,dt = 1$
- $F(t) = \text{Prob}[\ X \leq t\ ] = \int_{-\infty}^{t} f(s)\,ds.$

The expected value, also called the mean value, of X is

$$E(X) = \int_{-\infty}^{+\infty} t\,f(t)\,dt = \frac{1}{r}$$

For transfers from compartment I to$R$, the residence time is the period between time of infection and time of recovery, which is the infectious period. Then $\frac{1}{\gamma}$ is the mean infectious period.

Chapter 2

Some Classical Examples

2.1 Introduction

Ronald Ross discovers the transmission of Malaria by mosquitoes *Anopheles*. Ross proves this transmission in 1898 and was awarded Nobel prize in 1902. Ross is better known to the medical community as the discoverer of the mosquito transmission of malaria than as the author of a far-reaching theoretical approach to the study of disease in populations. We need little wonder that towards the end of his life, Ronald Ross, the man who incriminated the mosquito in the transmission of malaria, would write:

> 'In my own opinion my principal work has been to establish the general laws of epidemics'

This section will present the so-called Ross model. This model has been published in the appendix of in 1911 in *Prevention of Malaria* [78] and also in Nature the same year [79]. This model is interesting since

- It shows how to model, what are the hypotheses taken into account and the hypotheses neglected.

- this model is really seminal : the so-called famous model of Lotka-Volterra, also quadratic equations in the plane, is dated of 1925;
- despite its simplicity it captures the dynamics of Malaria;
- this model was used by Ross to found the justification of anti-vectorial measures.

The rest of the chapter is organized as follows. In accordance with the *a priori* method first we will describe the natural history of malaria. In other words, how does it work? Then we will discuss in detail the construction of Ross model and continue with his analytic study to state what Ross called the mosquito theorem (which also shows Ross state of mind). To conclude on malaria we will present a simple model intra-host, ie a model that describes the infection in an individual.

2.2 Natural history of Malaria

To be established Malaria needs three ingredients :

- A human host
- a mosquito of *Anopheles* type
- a hematophagous protozoan.

The causative agent, le *Plasmodium* was discovered in 1880, in Alg?ria at Constantine, by a french military doctor Alphonse Laveran. Laveran was awarded in 1907 by the Nobel prize.
4 parasitic species for man :

- *Plasmodium vivax*,

- *Plasmodium falciparum*,
- *Plasmodium malariae*,
- *Plasmodium ovale*.

All have an asexual cycle in man *s*chizogony and a sexual cycle in mosquito called *s*porogony. The most dangerous and frequent in Africa is *P. falciparum*. Parasite cycle is divided in 3 parts. Two in man and one in mosquito. The first part occurs in liver, the second occurs in red blood cells

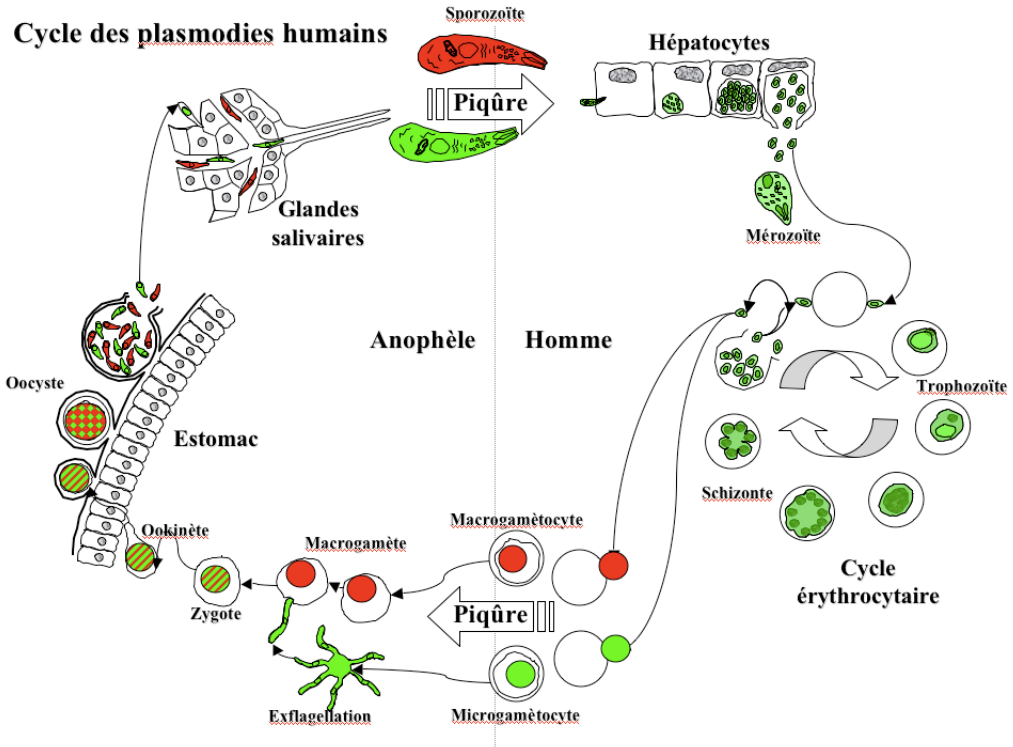


Figure 2.1: Le cycle du parasite du paludisme *Plasmodium falciparum* (D'apr?s C. Rogier)

2.2.1 In Liver

Parasite is inoculated in the peripheral blood with the anticoagulant saliva of moquito. . These parasites are called *sporozoites*. There are mobile and are moved by the blood flow to penetrate liver cells. It takes less than 45 minutes

Each sporozoite enters in a hepatocyte (liver cell); and from these moment unable to move Sporozoite transforms, grews and divides. After a mean duration between 8 to 15 days the hepatocyte is invaded by several thousands of nucleus called schizonte. Once matured the schizonte bursts and releases *m*?rozoites which pass in blood. The duration of this period is around 15 days

This Liver cycle was only discovered in 1948 seulement with teh works pf James, Tate, Shortt and Garnham.

2.2.2 In blood

M?rozoites invade Erythrocytes (Red Blood Cell) taking the characteristic aspect for *P. falciparum* of a kitten ring. They becomes *trophozoites* feeding from hemoglobin. At the end trophozoites becomes pigmented schizontes., Once mature the red blood cell burst and releases new merozoites which will parasite healthy RBC.

Several similar evolutions succeed one another as well. After few weeks (10-12 days for *P. falciparum*) some schizonts will turn into male sexed cells or females: the *g*ametocytes. We distinguish macrogametocytes (females) et des microgametocytes (males). This cells stays in the blood being a reservoir for the mosquito

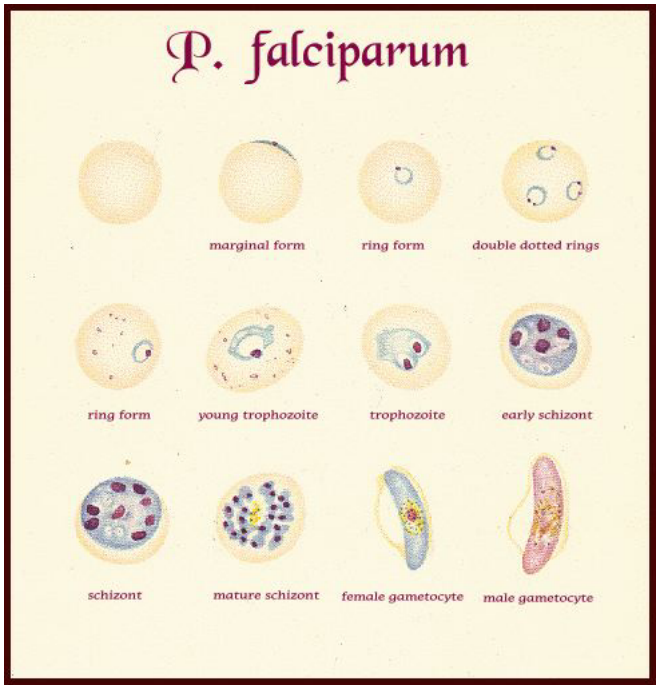


Figure 2.2: Different form of *Plasmodium falciparum* in Erythrocyte cycle (Laveran Drawings).

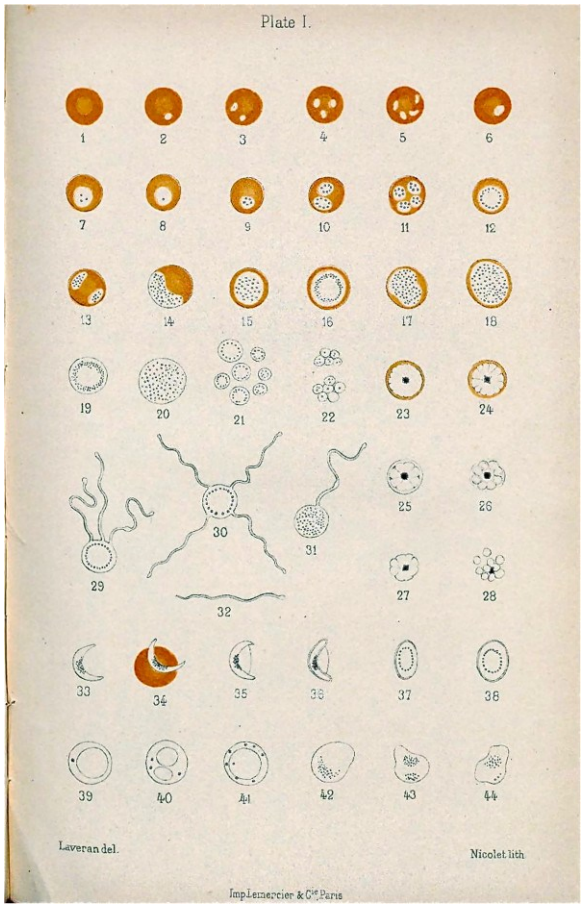


Figure 2.3: Laveran Drawings planche I [62]

sporogony in mosquito

If a female mosquito bites an infected individual it ingests gametocytes in his gut. The gametes increase in size. In 10 minutes the male and female gametes have left their envelope. The male produced 8 flagellated microgametes mobile. This event, the formation of vigorous mobile male gametes from a previously quiet gamete is called exflagellation. This striking phenomenon has fascinated malariologists since the observation by Laveran himself.

Fertilization produces a mobile ookinete that will establish itself as oocyst on the inside of the digestive tract.

Inside the oocyst will form sporocysts that will give several hundred sporozoites. The sporozoites migrate to the salivary glands of the mosquito, where they develop in vacuoles and can stay up to 59 days. During their development, sporozoites can become up to 1000 times more infectious than when their presence in the oocyst.

2.2.3 The vector

Only the female bite the host, usually after sunset. A blood meal is necessary before laying eggs singly on liquid surfaces.

Figure 2.4: anopheles

Eggs give larvae (Figure 2.5), then nymph and finally the winged insect

Figure 2.5: larvae ias parallel to surface

The flight of the mosquito does not exceed, in principle, one or two kilometers.

There are more than 300 species of Anopheles, only 60 are human plasmodium vectors.

2.3 Building the model

With Ross, we will make some simplifying hypotheses. It is assumed that the human population is constant as well as that of female anopheles. In other words, mortality is equal to the birth rate. A hypothesis of homogeneity is admitted: that humans and mosquitoes are equally distributed. In other words, a mosquito has an equal probability of biting a determined human.

The mosquito population is divided into two fictitious compartments: healthy mosquitoes, we say susceptible, and infectious mosquitoes. We do the same for the human population.

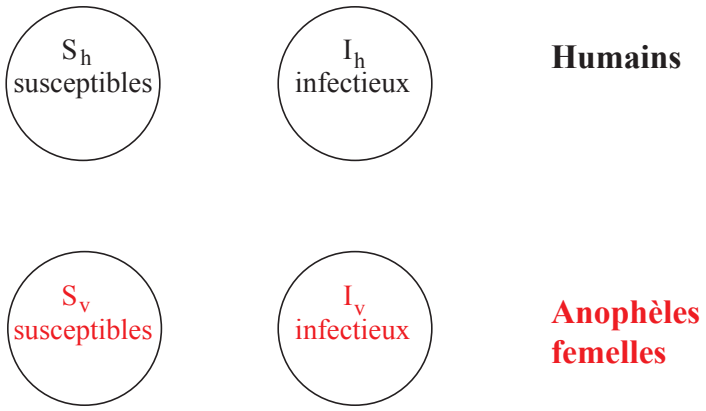


Figure 2.6: Les compartiments

These assumptions are simplifications. On the episode on which we study the transmission we can admit that the populations are approximately constant. In any model, there are simplifying hypotheses, e.g. when we write the equations of the pendulum we neglect the friction and the resistance of the air.

$S_h(t)$ and $I_h(t)$ are the respective populations of humans in the susceptible and infectious compartment.

We will write the balance of transfers between each compartment. We consider a time interval Δt, supposedly small. In this interval of time we will write the movements of populations between each compartment. There are some hidden hypotheses here: we neglect the incubation time, we also make the implicit assumption that there is no superinfection.

2.3.1 Infectious human evolution

W evaluate $I_h(t + \Delta t)$.

Input These are the new infectious individuals.

- To become infectious you must have been bitten by an infectious mosquito.
- A mosquito bites a human per unit of time.
- It is assumed that the probability of becoming infectious after an infectious bite is b_1.
- There are $I_v(t)$ infectious mosquitoes, they will induce $a\, I_v\, \Delta t$ bites.
- In all these bites, only those made on a susceptible human will produce a new infectious. The proportion is $\frac{S_h}{H} = \frac{H - I_h}{H}$ where H is the constant human population.
- Therefore the number of new infectious is

$$b_1\, a\, I_v\, \frac{H - I_h}{H}\, \Delta t$$

Output It's the infectious ones that heal and regain the susceptible compartment. It is therefore assumed that there is no immunity.

- It is assumed that the average speed for a healing individual is γ_H per unit of time. Mortality is assumed to be μ_H. This is the number of deaths per person per unit of time.
- Therefore it disappears

  $(\gamma_H + \mu_H)\, I_h\, \Delta t$

  infectious either by cure or by death.

Balance Finally

$$I_h(t+\Delta t) = I_h(t) + b_1\, a\, I_v\, \frac{H - I_h}{H}\, \Delta t - (\gamma_H\ + \mu_H)\, I_h\, \Delta t$$

Susceptibles balance Since the human population is constant , $S_h(t) = H - I_h(t)$ we have

$$S_h(t+\Delta(t) = S_h(t) - b_1\, a\, I_v\, \frac{H - I_h}{H}\, \Delta t + (\gamma_H\ + \mu_H)\, I_h\, \Delta t$$

Actually it is sufficient to know $I_h(t)$ to immediately know by difference $S_h(t)$. The first relation can also be written

$$\frac{I_h(t+\Delta t) - I_h(t)}{\Delta t} = b_1\, a\, I_v\, \frac{H - I_v}{H} - (\gamma_H\ + \mu_H)\, I_h$$

When Δt goes to 0 we obtain the following ODE

$$\frac{d\, I_h(t)}{dt} = \dot{I}_h(t) = b_1\, a\, I_v(t)\, \frac{H - I_h(t)}{H} - (\gamma_H\ + \mu_H)\, I_h(t)$$

which is written writes more simply, omitting the time t in the functions I_h and I_v.

$$\dot{I}_h = b_1\, a\, I_v\, \frac{H - I_h}{H} - (\gamma_H\ + \mu_H)\, I_h \tag{2.1}$$

This gives the following graph

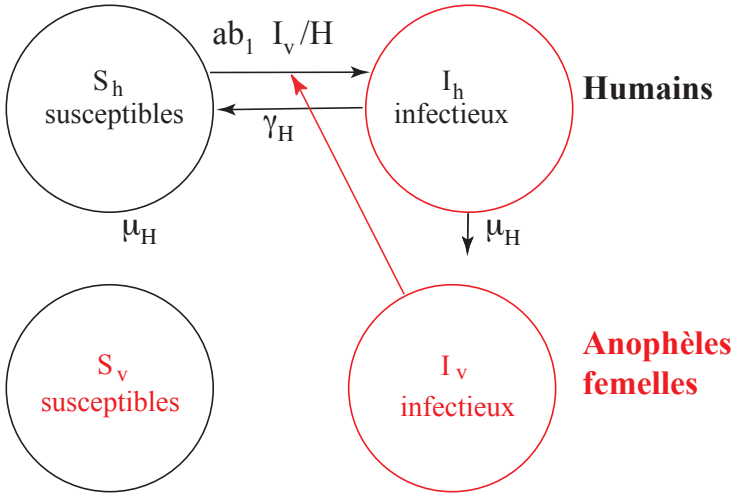


Figure 2.7: Human

2.3.2 Infectious mosquito population

The principle is the same. A new infectious mosquito will appear after the bite of a susceptible mosquito biting an infectious human. The probability of becoming infected, for the mosquito biting an infectious host, is b_2. We will have $a\,S_v$ bites, where $a\;S_v\,\frac{I_h}{H}$ will give rise to an infectious mosquito. If we denote by V the vector population (Anopheles females) $S_v = V - I_v$. We will thus have, by introducing the speed of recovering of the mosquito and its mortality

$$\dot{I}_v = b_2\,a\,(V - I_v)\,\frac{I_h}{H} - (\gamma_V + \mu_V)\,I_v \tag{2.2}$$

which gives the flow graph

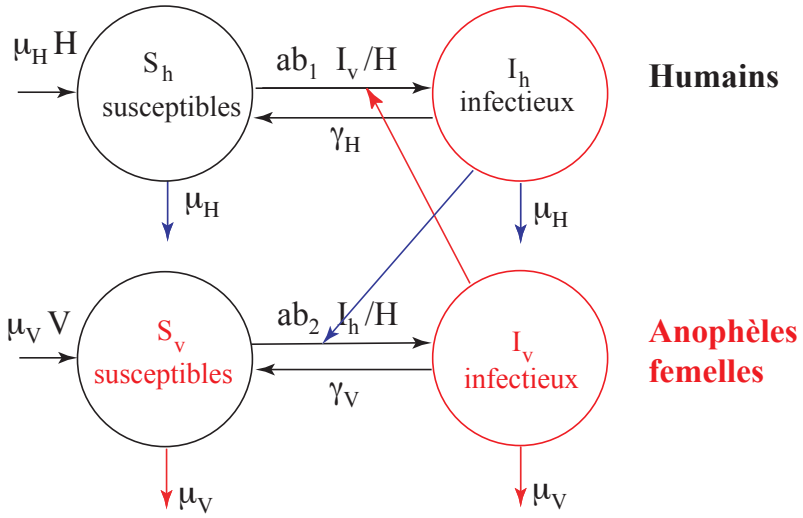


Figure 2.8: Complete flow graph

We see that in the human susceptible compartment, deaths are $\mu_H\, S_h$ and births $\mu_H\, H = \mu_H\, (S_h + I_h)$. Newborns are born susceptible. It is also an implicit assumption and it is actually true. The gain in the compartment of susceptible is in fact $\mu_H\, I_h$, in other words $\dot{S}_h = -\dot{I}_h$. Which is another way of saying that the H population is constant: $\dot{H} = 0$.

2.3.3 Ross model, final form

We have a system of ODE

$$\begin{cases} \dot{I}_h = & b_1\, a\, I_v\, \dfrac{H - I_h}{H} - (\gamma_H\, + \mu_H)\, I_h \\ \\ \dot{I}_v = & b_2\, a\, (V - I_v)\, \dfrac{I_h}{H} - (\gamma_V + \mu_V)\, I_v \end{cases} \tag{2.3}$$

In epidemiology, it is often the percentages, in other words the prevalences, that are measured. As the population is constant we will introduce the percentage of infectious individuals:

$$x = \frac{I_h}{H} \quad \text{for human hosts and} \qquad y = \frac{I_v}{V} \quad \text{for mosquitoes}$$

Since H et V are constant we have $\dot{x} = \dfrac{\dot{I}_h}{H}$ et $\dot{y} = \frac{\dot{I}_v}{V}$. We prepare, a litlle bit (2.3)

$$\begin{cases} \dot{I}_h = & b_1\, a\, \dfrac{I_v}{V}\, V\, \left(1 - \dfrac{I_h}{H}\right) - (\gamma_H\, + \mu_H)\, I_h \\ \\ \dot{I}_v = & b_2\, a\, V\, (1 - \dfrac{I_v}{V})\, \dfrac{I_h}{H} - (\gamma_V + \mu_V)\, I_h \end{cases} \tag{2.4}$$

dividing the first equation by H and the second by V, by setting $m = \frac{V}{H}$ we get

$$\begin{cases} \dot{x} = & m\, a\, b_1\, y\, (1 - x) - (\gamma_H\, + \mu_H)\, x \\ \\ \dot{y} = & b_2\, a\, (1 - y)\, x - (\gamma_V + \mu_V)\, y \end{cases} \tag{2.5}$$

To obtain the final model, two more approximations are made: the rate of recovering is the inverse of the average duration of the time spent in the infectious status. In other words, an infectious individual remains infectious on a mean time

$$\frac{1}{\gamma} \quad \text{time units}$$

In particular mortality is negligible in the face of healing time in humans. If conservatively we take between 2 and 6 months for a recovery

$\mu_H \approx 1/(60 \times 365)$ $j^{-1} \approx 4.56$ 10^{-5}

et $\gamma_V \approx 1/(2 \times 60)$ $j^{-}1 \approx 0.008$

$$\frac{\mu_V}{\gamma_V} \approx \frac{1}{360} \approx 0.0027$$

Similarly, the mosquito's recovering time is assumed to be negligible compared to life expectancy of the mosquito. In all the entomological literature it is admitted that the mosquito remains infected all his life. We therefore neglect μ_H and γ_V. Which finally gives

$$\begin{cases} \dot{x} = & m\,a\,b_1\,y\,(1-x) - \gamma\,x \\ \\ \dot{y} = & b_2\,a\,(1-y)\,x - \mu\,y \end{cases} \tag{2.6}$$

As a result of this model Ross stated what he called his mosquito theorem. Formulated in a contemporary way, this theorem would be read now

Theorem 2.3.1 (Mosquito theorem, Ross)
For the system (2.6)

- *Si on a* $\dfrac{m\,a^2\,b_1\,b_2}{\gamma\,\mu} \leq 1$,
  the disease free equilibrium $(0,0)$ *is globally asymptotically stable on* $[0,1]\times[0,1]$.
- *Si* $\dfrac{m\,a^2\,b_1\,b_2}{\gamma\,\mu} > 1$,
  then there exists a unique equilbrium $(\bar{x},\bar{y}) \in]0,1[\times]0,1[\ \ ]0,1]\times]0,1]$ *which is is globally asymptotically stable on* $]0,1[\times]0,1[$.

2.4 Ross model analysis

We will postpone this analysis, waiting for the tools needed.

2.5 Malaria intra-host model

Ross model is a model for the spread of a disease. It may be useful to study the spread of pathogens within an individual. The model we will present was introduced by Anderson, May and Gupta in 1989.

Let x the concentration in the blood of healthy erythrocytes, y the concentration of parasitized erythrocytes and m the concentration of merozoites circulating freely in the bloodstream.

$$\begin{cases} \dot{x} = \Lambda - \mu_x\, x - \beta x\, m \\ \\ \dot{y} = \beta\, x\, m - \mu_y\, y \\ \\ \dot{m} = r\, \mu_y\, y - \mu_m\, m - \beta\, x\, m. \end{cases}$$

In the absence of parasites, the concentration of red blood cells is constant. The number of red blood cells in the blood is normally between 4.5 - 5.5 million / mm3, their lifespan is 120 days, they are produced by the bone marrow. This explains the choice of $\dot{x} = \Lambda - \mu_x\, x$ in the absence of parasite. Now the term $\beta\, x\, m$ represents the penetration rate of merozoites in erythrocytes An infected erythrocyte passes into the infected compartment The mortality of infected erythrocytes is μ_y. For *P. falciparum* the average cycle time is 48 hours When a red blood cell bursts it releases r merozoites. A merozoite if it does succeed to enter a RBC will be eliminated in the spleen. The term $\beta\, x\, m$ which appears in the last equation represents the passage of the merozoite circulating in the red blood cell. If this term was not present, the model would allow a single merozoite to infect several RBC!

The term $r\, \mu_y$ is the mean number of sporozoites produced by a infected erythrocyte by unit of time. Since an erythrocyte has a mean life of $\dfrac{1}{\mu_y}$ and since when bursting it gives r merozoites, the number by unit of time is

$$\frac{r}{\frac{1}{\mu_y}} = r\,\mu_y$$

In this model the unknown parameter is β, for the others we have at least an approximate knowledge.

This model, without the quadratic term in $\dot{m}$, is baptized as a model of viral dynamics (May et al)

In fact this model is also a model of HIV infection. It is proposed by Perelson in the 1990s. The only difference is that recruitment, instead of being constant is represented by a logistic function. In this case, the variable x represents the concentration of CD4 $^+$ lymphocyte cells. and we would have

$$\dot{x} = \Lambda + p\,x\,\left(1 - \frac{x}{x_{\max}}\right) - \mu_x\,x - \beta\,x\,m.$$

In this form y is the concentration of infected lymphocytes and m the concentration of free circulating virion.

With this model Perelson has said

> While the mathematics involved was trivial, the application of mathematics in this manner was novel and set off what has been described as a revolution in thinking about HIV

2.6 $SEIR$ model

Disease infection begins with the transmission of the pathogen from one host to another. After pathogens invade the host body, they need to be able to evade or overcome the host immune response, and be able to multiply or replicate. When the pathogens accumulate sufficiently large numbers and reach the targeted organs, they begin to cause sufficient damage to the host

body so that the host becomes symptomatic, and the host is then capable to transmit the pathogens to others. The period from time of infection to time of onset of symptoms is called the incubation period. The period from time of infection to time of being contagious or infectious is called the latent period. The period from the beginning to the end of being infectious is called the infectious period. See the illustration (2.9) for an example of relations between these periods. During the latent period, a host may or may not show symptoms, but the host is not capable of transmitting pathogens to other hosts.

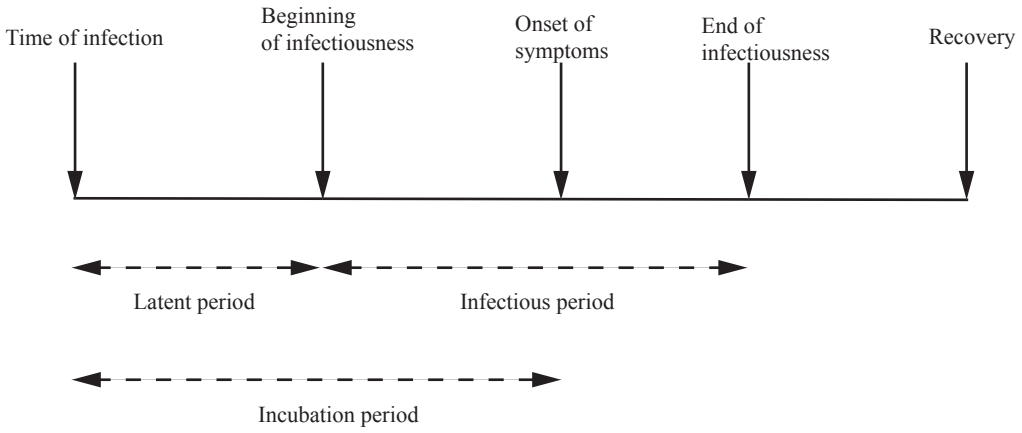


Figure 2.9: SEIR model

We have the following flow graph

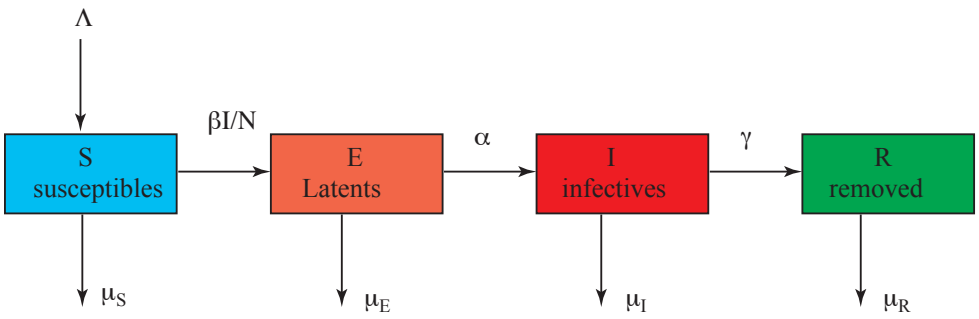


Figure 2.10: Scheme of infection

$$\begin{cases} \dot{S} = & \Lambda - \mu_S\, S - \beta\, \dfrac{S\, I}{N} \\ \dot{E} = & \beta\, I\, S - (\mu_E + \alpha)\, E \\ \dot{I} = & \alpha\, I - (\gamma + \mu_I)\, I \\ \dot{R} = & \gamma\, I - \mu_R\, R \end{cases} \tag{2.7}$$

In this model S, E, I, R and $N = S+E+I+R$ are numbers. It is assumed, under the hypothesis of homogeneity that any individual makes β adequate contact by unit of time. Then I infectious will make $\beta\, I$ adequate contacts. But in all these contacts, only the contact with susceptible individuals will give rise to latent (E) individual. The proportion of susceptibles in the whole population is $\dfrac{S}{N}$. Then the new latent will be

$$\beta\, I \frac{S}{N}$$

This law is called true mass action or frequency-dependent transmission. The mean period for latency is $\dfrac{1}{\alpha}$, for recovery $\dfrac{1}{\gamma}$

Chapter 3

Basic Mathematical Tools and Techniques

3.1 Well-posedness of a model

We consider the Kermack-McKendrick model (1.1)

$$\begin{cases} \dot{S} &= -\beta\, S\, I \\ \dot{I} &= \beta\, S\, I - \gamma\, I \\ \dot{R} &= \gamma\, I \end{cases} \tag{3.1}$$

with initial condition (S_0, I_0, R_0). We claim that nonnegative initial conditions leads to nonnegative solutions. In other word, any trajectory beginning in the nonnegative orthant $\mathbb{R}^3_+$ stays in this orthant. In other words the nonnegative orthant is positively invariant for the dynamical system. It is what we mean by well-posedness. Recall that the variables are either numbers or densities, then in the nonnegative orthant.

To study the well-posedness of a system we will give a Theorem. This Theorem seems to be intuitively evident, but it has to be proved ...

This Theorem will allows to study well-posedness for epridemiological or

biological models.

Theorem 3.1.1 (Barrier Theorem)

We consider a differential equation $\dot{x} = X(x)$, where X is a $\mathcal{C}^1$ function defined on an open set $U \subset \mathbb{R}^n$, $U \longrightarrow \mathbb{R}^n$.
We consider a $\mathcal{C}^1$ function $H : \mathbb{R}^n \longrightarrow \mathbb{R}$. We define a closed set

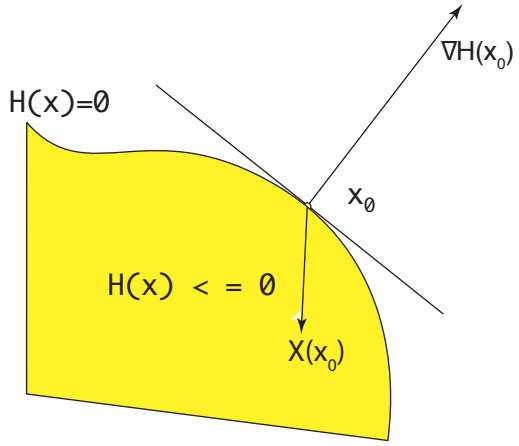
$$\Omega = \{x \in \mathbb{R}^n \mid H(x) \leq 0\}$$

and its boundary is $\partial\Omega = \{x \in \mathbb{R}^n \mid H(x) = 0\}$.

We assume that in every point of the boundary $x \in \partial\Omega$ we have $\nabla H(x) \neq 0$ and

$$\langle X(x) \mid \nabla H(x) \rangle \leq 0,$$

then the set Ω is positively invariant.



In other words $\{x \in \mathbb{R}^n \mid H(x) \leq 0 \quad x \in \overline{\Omega}\}$ is positively invariant.This result seems to be intuitively evident. This partly true as the proof will show.

Geometrically this says that the vector field, on the boundary 'points inside'.

before proving the Theorem we need another result. This result tell how we can have estimates when we modify (approximate) a vector field

Lemma 3.1.1

Let X a Lipschitz vector vector with Lipschitz constant L. We consider the approximation X^ε of X, in other words for any x we have

$$\|X^\varepsilon(x) - X(x)\| \leq \varepsilon$$

$\|\ \|$ *being any norm on* $\mathbb{R}^n$.
Then for any t where the quantities are defined

$$\|X_t^\varepsilon(x_0^\varepsilon) - X_t(x_0)\| \leq \|x_0^\varepsilon - x_0\|\, e^{Lt} + \varepsilon\, \frac{e^{Lt} - 1}{L}$$

proof of the Theorem

Actually we will prove the theorem for a Locally Lipschitz vector field. The case $\mathcal{C}^1$ is then contained in it. To go out from the set

$$G = \{x \in \mathbb{R}^n \mid H(x) \leq 0 \ \ x \in \overline{\Omega}\}$$

the trajectory, by the intermediate value theorem, must pass through he boundary $\partial G = \{x \in \overline{\Omega} \mid H(x) = 0 \ \}$
We will distinguish two cases.

In the first case

we suppose that in $x_0, H(x_0) = 0$. We have

$$\langle X(x_0) \mid \nabla H(x_0)\rangle < 0$$

Let $\varepsilon < 0$ such that $\langle X(x_0) \mid \nabla H(x_0)\rangle < \varepsilon < 0$, with a continuity argument, there is a ball centered in x_0 and of radius $\eta > 0$, such that for all $y \in B(x_0, \eta)$ we have

$$\langle X(y) \mid \nabla H(y) \rangle < \varepsilon < 0$$

We consider the trajectory $X_t(x_0)$ from x_0. For $t \geq 0$ small enough $0 \leq t < \alpha$, the trajectory remains in the ball $B(x_0, \eta)$. We have

$$\frac{d}{dt} H(X_t(x_0)) = \langle \nabla H(X_t(x_0)) \mid X\left(X_t(x_0)\right) \rangle < \varepsilon < 0$$

The function $H(X_t(x_0))$ is strictly decreasing and so $H(X_t(x_0)) < 0$ for $0 < t < \alpha$.

Which proves that $X_t(x_0) \in \overset{\circ}{G}$

Second case

we suppose now that $\langle X(x_0) \mid \nabla H(x_0) \rangle = 0$. We consider the vector field

$$X^{\varepsilon}(x) = X(x) - \varepsilon \frac{\nabla H(x)}{\|\nabla H(x)\|}$$

This vector field satisfies for all $\varepsilon > 0$, the hypothesis of first case on $\Omega \cup \partial G$. Let η small enough such that in the closed ball $B(x_0, \eta)$ the vector field X^{ε} satisfies the required inequality. We choose $t \leq T$ sufficiently small such that $X_t(x_0) \in B(x_0, \eta/2)$. Since X^{ε} is a ε approximate field of X, we apply the lemma (3.1.1), which gives us the increase

$$\|X_t^{\varepsilon}(x_0) - X_t(x_0)\| \leq \varepsilon \, \frac{e^{L\,T} - 1}{L}$$

This proves that by choosing ε small enough such that $\frac{e^{L\,T}-1}{L} < \eta/2$ we will have

$$X_t^{\varepsilon}(x_0) \in B(x_0, \eta)$$

From the previous demonstration $X_t^\varepsilon(x_0) \in \overset{\circ}{G}$, so $X_t(x_0)$ is the limit of points of G which is closed, so in G. The path from x_0 can not leave G locally. Since this is true for every x_0 point of ∂G we have shown the result on Ω.

■

Proof of the lemma

We use the fondamental identity

$$X_t(x_0) = x_0 + \int_0^t X\left(X_s(x_0)\right)\, ds$$

then

$$\|X_t^\varepsilon(x_0^\varepsilon) - X_t(x_0)\| \leq \|x_0^\varepsilon - x_0\| + \int_0^t \|X^\varepsilon\left(X_s^\varepsilon(x_0^\varepsilon)\right) - X((X_s(x_0))\,\|ds$$

Writting

$$\int_0^t \|X^\varepsilon\left(X_s^\varepsilon(x_0^\varepsilon)\right) - X((X_s(x_0))\,\|ds \leq$$
$$\int_0^t \|X^\varepsilon\left(X_t^\varepsilon(x_0^\varepsilon)\right) - X((X_s^\varepsilon(x_0))\,\|ds + \int_0^t \|X\left(X_t^\varepsilon(x_0^\varepsilon)\right) - X((X_s(x_0))\,\|ds$$

we get

$$\|X_t^\varepsilon(x_0^\varepsilon) - X_t(x_0)\| \leq \|x_0^\varepsilon - x_0\| + \varepsilon\, t + L\int_0^t \|X_s^\varepsilon(x_0^\varepsilon) - X_s(x_0)\|\, ds$$

If we set $u(t) = \|X_t^\varepsilon(x_0^\varepsilon) - X_t(x_0)\|$ we have

$$u(t) \leq u(0) + \varepsilon\, t + L\int_0^t u(s)\, ds$$

The inequality is immediate by Gronwall's lemma.

■

Lemma 3.1.2 (Gronwall 2) *Let $t\ u : [0, \alpha] \longrightarrow \mathbb{R}^+$ a continuous and nonnegative function. We assume that there exists constants C, ε and L such that for any $t \in [0, \alpha]$ we have*

$$u(t) \leq C + \varepsilon\, t + L \int_0^t u(s)\, ds \tag{3.2}$$

Then we have

$$u(t) \leq C\, e^{Lt} + \frac{\varepsilon}{L}(e^{Lt} - 1)$$

Proof

Let $y(t) = L \displaystyle\int_0^t u(s)\, ds$ et $z(t) = e^{-Lt}\, y(t)$. From inequality 3.2.

$$\dot{y} \leq L\,(\,C + \varepsilon\, t) \;+\; L\, y$$

and

$$\dot{z} = e^{-Lt}(\dot{y} - L\, y) \leq L e^{-Lt}(C + \varepsilon t)$$

which gives integrating from 0 to t

or equivalently

$$z(t) \leq \int_0^t L\, e^{-Ls}(C + \varepsilon s)\, ds$$

$$y(t) \leq e^{Lt} \int_0^t L\, e^{-Ls}(C + \varepsilon s)\, ds$$

taking into account the inequality satisfied by u

$$u(t) \leq C + \varepsilon t + e^{Lt} \int_0^t L\, e^{-Ls}(C + \varepsilon s)\, ds$$

A straighforward computation leads to

$$e^{Lt}\int_0^t L\,e^{-Ls}(C+\varepsilon s)\,ds = -C-\varepsilon t + Ce^{Lt} + \frac{\varepsilon}{L}(e^{Lt}-1)$$

Which is the inequality to prove ■

Remark 3.1.1

If the vector field is not Lipschitzian, the theorem is no more true. For example for the equation $\dot{x} = -3\,|x|^{\frac{2}{3}}$ *on* $\mathbb{R}$*, origine is a barrier however some solutions can go through. For this vector fiels* $\mathbb{R}+$ *is not positively invariant.*

The same applies to the vector field given by the ODE $\dot{x} = -\sqrt{2\,g\,x}$*, the origin is a barrier and* $\mathbb{R}_-$ *should be positively invariant. This is not the case as can be seen on the figure 3.1.*

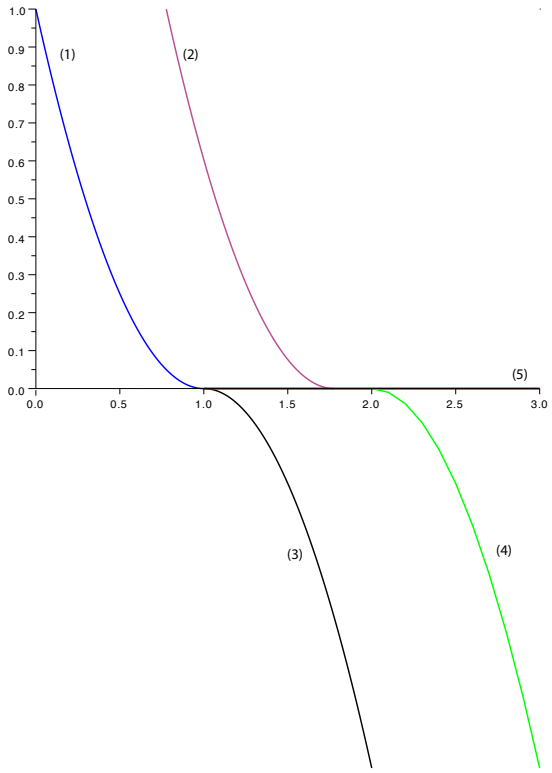


Figure 3.1: Different solutions of $\dot{x} = -\sqrt{2\,g\,x}$

3.1.1 Examples

3.1.2 Kermack-McKendrick model

$$\begin{cases} \dot{S} &= -\beta\, S\, I \\ \dot{I} &= \beta\, S\, I - \gamma\, I \\ \dot{R} &= \gamma\, I \end{cases}$$

We prove that the nonnegative orthant is positively invariant. The boundary of $\mathbb{R}^3_+$ is given by 3 hyperplane cone : $\mathbb{R}_+ \times \mathbb{R}_+ \times \{0\}$, $\{0\} \times \mathbb{R}_+ \times \mathbb{R}_+$

and $\mathbb{R}_+ \times \{0\} \times \mathbb{R}_+$.

When $S = 0$ we have $\dot{S} = 0$, the vector field is tangent to the RI-plane hence the conditions of theTheorem are satisfied. Identically if $I = 0$ then $\dot{I} = 0$. When $R = 0$, $\dot{R} = \gamma\, I \geq 0$, since we are in the nonnegative orthant. Note that to be in accordance with the notations of the Theorem, the nonnegative orthant is defined by $-S \leq 0\,; -I \leq 0\,; -R \leq 0$.

3.1.3 Ross model

3.2 Lyapunov techniques

Lyapunov method also called direct method or second method of Lyapunov has been introduced in 1892 in Lyapunov's thesis and published in french in 1897.

Probl?me g?n?ral de la stabilit? du mouvement. Annales de la facult? des sciences de Toulouse 9(2) (1907): 203-474

This method allow to establish stability of the equilibrium of a system, without integrate this EDO.

This method was forgotten and rediscovered in USSR in the 1944

This method was ignored in the West till 1950. From this date, this method was the prerogative of the russian mathematician and control engineers. Its importance was rediscovered first in control theory and popularized by LaSalle and Lefschetz in 1959.

It is now well funded that Lyapunov is a very fundamental method.

3.2.1 Problematics

$$\left\{\begin{array}{rl} \dot{x} & = f(x) \\ & x \in \mathbb{R}^n \\ x(0) & = x_0 \end{array}\right. \tag{3.3}$$

The unique solution for the initial value x_0, is denoted by $\Phi_t(x_0)$. By renormalizing f, we can always suppose that the vector field is complete, i.e., the function $\Phi_t(x_0)$ is defined for any t.

Now we suppose that x_0 is an equilibrium,

$$f(x_0) = 0$$

We have the following well known property : for any $(t, s) \in \mathbb{R}^2$

$$\Phi_t\left(\Phi_s(x)\right) = \Phi_{t+s}(x)$$

One parameter group property

We recall properties of equilibria

Definition 3.2.1 (stability) *We say that x_0, an equilibrium of $\dot{x} = f(x)$ is stable (in Lyapunov's sense) iff for any open set U containing x_0, it exists an open set V of initial values , $V \subset U$ such that for any $y \in V$ and for any $t \geq 0$ we have $\Phi_t(y) \in U$*

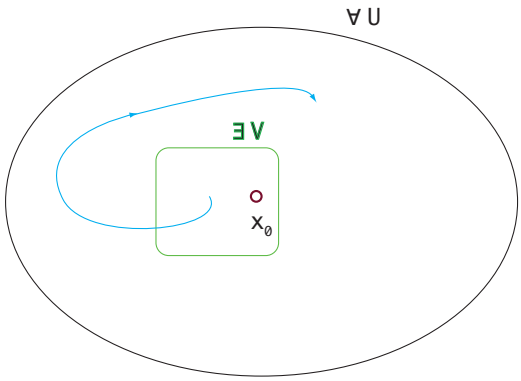


Figure 3.2: Stable equilibrium

Definition 3.2.2 (attractivity) *We say that x_0 is attractive in the open set V if for any $y \in V$*

$$\lim_{t \to +\infty} \Phi_t(y) = x_0$$

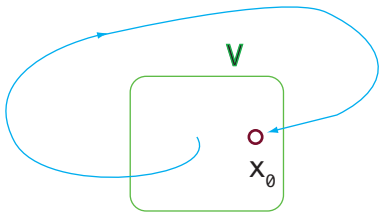


Figure 3.3: attractive equilibrium

Definition 3.2.3 (asymptotic stability) *We say that x_0 is asymptotically stable (locallly) if x_0 is stable and if there exists an open set V of x_0 in which x_0 is attractive.*

Remark 3.2.1 *Beware : attractivity does not imply stability. However this is true for linear systems. $\dot{x} = A\,x$.*

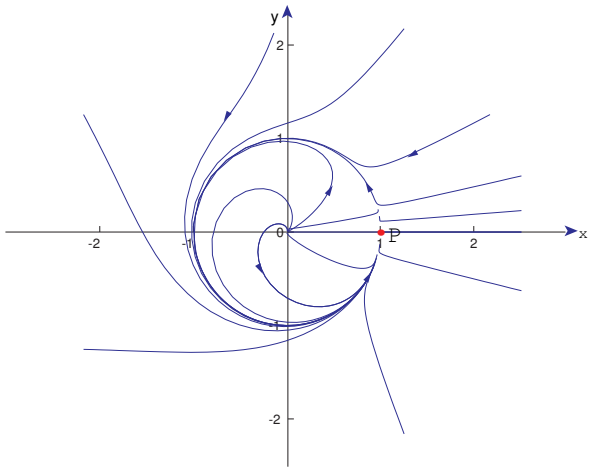


Figure 3.4: attractivity without stability

3.2.2 Lyapunov functions

Definition 3.2.4 (Lyapunov fonction) *We call Lyapunov function in x_0, an equilibrium of $\dot{x} = f(x)$, a function V such that there is an open set U containing x_0, and such that the following properties are satisfied*

- $V(x) \geq 0$ *sur* U
- $V(x) = 0$ *iff* $x = x_0$
- *On* U *we have*

$$\dot{V}(x) = \langle \nabla V(x) \mid f(x) \rangle \leq 0$$

A function satisfying the two first properties in x_0 is said definite positive.

3.2.3 Theorems

Theorem 3.2.1 (Lyapunov first theorem) *If x_0 is an equilibrium of $\dot{x} = f(x)$, if there exists a Lyapunov function in x_0 for this system then x_0 is a stable*

equilibrium.

Lyapunov second theorem

If moreover $\dot{V}$ is negative definite , i.e. si $\dot{V}(x) = 0$ iff $x = x_0$, then x_0 is an asymptotically stable equilibrium

The attraction basin of x_0 is contained in U where the 3 properties of V are satisfied.

Theorem 3.2.2 (Lasalle) *If V is a Lyapunov function the greatest invariant set contained in*

$$\mathcal{L} = \{x \mid \dot{V}(x) = 0\}$$

is an attractive set. This assertion is called LaSalle's principle of invariance and is true even is V is only nonnegative and non necessarily positive definite. If $\mathcal{L} = \{x_0\}$ then x_0 is asymptotically stable.

Theorem 3.2.3 (Lasalle) *We consider the ODE with an equilibrium x_0, defined on a compact positively invariant set Ω.*

If we have a nonnegative function V, such that $\dot{V} \leq 0$ on Ω and moreover the largest invariant set contained in $\mathcal{L} = \{x \in \Omega \mid \dot{V}(x) = 0\}$ is reduced to $\{x_0\}$

Then x_0 is globally asymptotically stable in Ω

Note: V is not a Lyapunov function : not positive definite.

Theorem 3.2.4 (Poincar?-Lyapunov) *We consider a $\mathcal{C}^1$ ODE , $\dot{x} = f(x)$ and x_0 and equilibrium.*

1. *If $Df(x_0)$ has all its eigenvalues with negative real part, i.e., $s(Df(x_0)) < 0$, then x_0 est asymptotically (locally) stable.*

2. *if* $Df(x_0)$ *has (at least) one eigenvalue with a positive real part, i.e.,* $s(Df(x_0)) > 0$*, then* x_0 *is unstable.*

Advantage of Lyapunov over Poincar? : Lyapunov can give a conclusion when Poincare fails

example :

$$\begin{cases} \dot{x} &= -x^2 \\ \dot{y} &= -y \end{cases}$$

The positive orthant is est positively invariant. The origin is stable in the domain $\mathbb{R}_+ \times \mathbb{R}_+$.

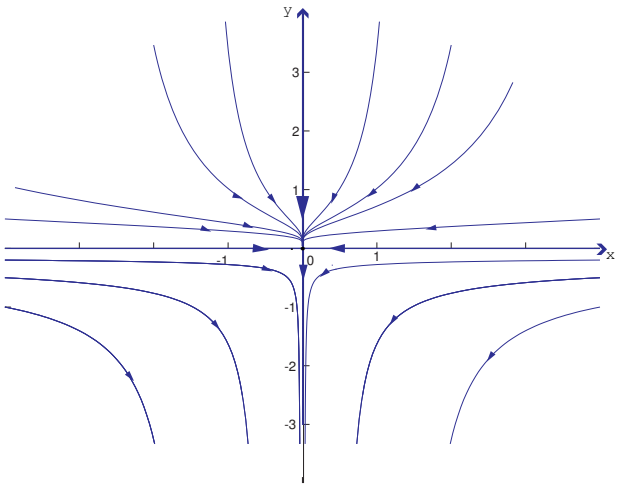


Figure 3.5: Saddle-Node

Il is sufficient to choose for Lyapunov function on sur $\mathbb{R}_+ \times \mathbb{R}_+$

$$V(x, y) = x + \frac{1}{2} y^2$$

3.2.4 Examples

Lotka-Volterra model introduced by Pi?lou : n the prey , p the predator

$$\begin{cases} \dot{n} &= r\,n\,\left(1-\frac{n}{K}\right) - a\,n\,p \\ \dot{p} &= b\,n\,p - \mu\,p \end{cases}$$

This can be also considered as an intra-host model for a disease. n can represent a concentrationof target cells, e.g. red blood cells and p a parasit destroying red blood cells...

It is easy to determine a coexistence equilibrium

$$n^* = \frac{\mu}{b} \quad p^* = \frac{r}{a}\left(1-\frac{\mu}{b\,K}\right)$$

This equilibrium has a biological meaning if $p^* > 0$ or $\frac{b\,K}{\mu} > 1$

This coefficient is a basic reproduction ratio : it is the mean number of predator fathered by one predator introduced in a population of preys (without predator) during its life :

Prey Population at Equilibrium : K , mean life of a predator $\frac{1}{\mu}$, basic repruction ration $b\,K\,\frac{1}{\mu}$.

Proof of the Stability of the coexistence equilibrium when $\mathcal{R}_0 = \frac{b\,K}{\mu} > 1$.

we consider

$$f(n,p) = b\,(n - n^*\,\ln n) + a\,(p - p^*\,\ln p)$$

and the Lyapunov function $\mathbb{R}_+^* \times \mathbb{R}_+^*$

$$V(n,p) = b\,(n - n^*\,\ln n) + a\,(p - p^*\,\ln p) - f(n^*,p^*)$$

$$\dot{V} = b\,r\,n\left(1-\tfrac{n}{K}\right) - a\,b\,n\,p - b\,r\,n^*\left(1-\tfrac{n}{K}\right) + a\,b\,n^*\,p + a\,b\,n\,p - a\,\mu\,p - a\,p^*\,(b\,n-\mu)$$

Taking into account $b\,n^* = \mu$ we get

$$\dot{V} = b\,r\,\left(1 - \frac{n}{K}\right)\,(n - n^*) - a\,b\,p^*\,(n - n^*)$$

Using again $a\,p^* = r\,\left(1 - \frac{n^*}{K}\right)$ we obtain

$$\dot{V} = b\,(n - n^*)\,r\,\left(1 - \frac{n}{K} - 1 + \frac{n^*}{K}\right) = -\frac{b\,r}{K}\,(n - n^*)^2 \leq 0$$

This proves stability . Now consider the set $\mathcal{L}$ defined by

$$\mathcal{L} = \{(n, p) \geq 0 \;\mid\; n = n^*\}$$

To be invariant in this set, n must be constant equal to n^* must be constant, then $\dot{n} = 0$. Hence

$$\dot{n} = r\,n^*\,\left(1 - \frac{n^*}{K}\right) - a\;n^*\,p = 0$$

Precisely $p = p^*$. The greatest invariant set $\mathcal{L}$ is $\{(n^*, p^*)\}$.

We conclude by LaSalle 's invariance principle to the asymptotic stability on $\mathbb{R}_+^* \times \mathbb{R}_+^*$.

Stability of the predator free equilibrium when $\mathcal{R}_0 = \frac{b\,K}{\mu} \leq 1$.

Lyapunov function

$$V(n, p) = b\,(n - K\;\ln n) + a\,p$$

Which gives

$$\dot{V} = -\frac{b\,r}{K}\,(n - K)^2 + a\,\mu\,p\,(\mathcal{R}_0 - 1) \leq 0$$

3.2.5 How to find a Lyapunov function ?

Bad news

More an art than a science

Good news :

Some ingenuity, astuteness and tricks are needed

How do you find this dawn Lyapunov function for the Lotka-Volterra Model ?

Back to classical Lotka-Volterra

$$\begin{cases} \dot{n} &= r\,n\, - a\; n\,p \\ \\ \dot{p} &= b\,n\,p - \mu\,p \end{cases}$$

Dividing the two equations

$$\frac{dn}{dp} = \frac{r\,n\, - a\; n\,p}{b\,n\,p - \mu\,p}$$

We can separate the variables

$$(b - \frac{\mu}{n})\,dn = (-a + \frac{r}{p})\,dp$$

Equivalently

$$(b - \frac{\mu}{n})\,dn - (-a + \frac{r}{p})\,dp = 0$$

Integrating this relation shows that

$$f(n,p) = b\,\left(n - \frac{\mu}{b}\,\ln n\right) + a\,\left(p - \frac{r}{a}\,\ln p\right)$$

is a first integral, i.e., the derivative of f on the trajectories are zero, or this function is constant on the trajectories of the ODE. Recall : the coexistence equilibrium is $n^* = \frac{\mu}{b}$, $p^* = \frac{r}{a}$

Interlude : the function $s - s^*\,\ln s$, defined on $\mathbb{R}^+\backslash 0$ has a unique minimum s^*. Hence

$$s - s^* \ln s - s^* + s^* \ln s^*$$

is definite positive for s^*

Recall

$$f(n,p) = b\left(n - \frac{\mu}{b}\ln n\right) + a\left(p - \frac{r}{a}\ln p\right)$$

Tada !!!

$$f(n,p) - f(n^*,p^*)$$

is a Lyapunov function for (n^*, p^*)

The function

$$f(n,p) = b\,(n - n^* \ln n) + a\,(p - p^* \ln p)$$

is a Lyapunov function for the classical Lotka-Volterra. For the Pielou Lotka-Volterra we use the same function, evidently with modified value for the equilibrium.

If $\frac{b\,K}{\mu} > 1$

$$n^* = \frac{\mu}{b} \quad p^* = \frac{r}{a}\left(1 - \frac{\mu}{b\,K}\right)$$

If

$\frac{b\,K}{\mu} < 1$

$n^* = K$ and $p^* = 0$ which makes disappear the $p^* \ln p$ term.

3.2.6 Lyapunov and Ross model

Theorem 3.2.5 *Let G an open set, containing origin, positively invariant for the system*

$$\dot{x} = A(x).x,$$

where $A(x)$ is a Metzler matrix depending continuously of x.

We suppose there exists $c^T \gg 0$ such that $c^T A(x) \ll 0$ for any $x \in G$, $x \neq 0$.

Then the origin is GAS in G.

Consider on G Lyapunov function.

$$V(x) = \sum_{i=1}^{n} c_i \mid x_i \mid .$$

We define $\varepsilon_z = \text{sign}(z)$, i.e. $|x_i| = \varepsilon_{x_i} x_i$.

The function V is locally Lipschitz : we can defined Dini derivative.

$$\begin{aligned}
\dot{V} &= \textstyle\sum_{i=1}^{n} c_i \varepsilon_{x_i} \dot{x}_i \\
&= \textstyle\sum_{i=1}^{n} c_i \varepsilon_{x_i} \sum_{j=1}^{n} a_{ij} x_j \\
&= \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} c_i \varepsilon_{x_i} a_{ij} x_j \\
&= \textstyle\sum_{j=1}^{n} \varepsilon_{x_j} x_j \sum_{i=1}^{n} c_i \varepsilon_{x_j} \varepsilon_{x_i} a_{ij} \\
&= \textstyle\sum_{j=1}^{n} \varepsilon_{x_j} x_j \left[c_j a_{jj} + \sum_{i \neq j} c_i \varepsilon_{x_j} \varepsilon_{x_i} a_{ij} \right] \\
&\leq \textstyle\sum_{j=1}^{n} \varepsilon_{x_j} x_j \left[c_j a_{jj} + \sum_{i \neq j} c_i a_{ij} \right] = \sum_{j=1}^{n} |x_j| \, (c^T A)_j \leq 0.
\end{aligned}$$

Since $c^T A(x) \ll 0$ sur G, function $\dot{V}$ is definite negative.

$$\begin{cases} \dot{x} = & m\, a\, b_1\, y\, (1-x) - \gamma\, x \\ \\ \dot{y} = & b_2\, a\, (1-y)\, x - \mu\, y \end{cases}$$

Two equilibria : DFE : $(0, 0)$ and

$$\bar{x} = \frac{\frac{m\,a^2\,b_1\,b_2}{\mu\,\gamma} - 1}{\frac{m\,a^2\,b_1\,b_2}{\mu\,\gamma} + \frac{b_2\,a}{\mu}} \qquad \bar{y} = \frac{\frac{m\,a^2\,b_1\,b_2}{\mu\,\gamma} - 1}{\frac{m\,a^2\,b_1\,b_2}{\mu\,\gamma} + \frac{m\,b_1\,a}{\gamma}}$$

This equilibrium has a biological meaning iff

$$\mathcal{R}_0 = \frac{m\,a^2\,b_1\,b_2}{\mu\,\gamma} > 1$$

$$\begin{cases} \dot{x} = & \alpha\,y\,(1-x) - \gamma\,x \\ \\ \dot{y} = & \beta\,(1-y)\,x - \mu\,y \end{cases}$$

Two equilibria (DFE) : $(0,0)$ and

$$\bar{x} = \frac{\frac{\alpha\,\beta}{\mu\,\gamma} - 1}{\frac{\alpha\beta}{\mu\,\gamma} + \frac{b_2\,a}{\mu}} \qquad \bar{y} = \frac{\frac{\alpha\,\beta}{\mu\,\gamma} - 1}{\frac{\alpha\,\beta}{\mu\,\gamma} + \frac{m\,b_1\,a}{\gamma}}$$

Make sense iff

$$\mathcal{R}_0 = \frac{\alpha\,\beta}{\mu\,\gamma} > 1$$

We can write

$$\begin{bmatrix} \dot{x} \\ \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\gamma & \alpha\,(1-x) \\ \\ \beta\,(1-y) & -\mu \end{bmatrix} \begin{bmatrix} x \\ \\ y \end{bmatrix}$$

Which is

$$\dot{X} = A(X)\,X$$

Stability of the DFE

$$\mathcal{R}_0 = \frac{\alpha\,\beta}{\mu\,\gamma} \leq 1$$

$$A(x,y) = \begin{bmatrix} -\gamma & \alpha\,(1-x) \\ \beta\,(1-y) & -\mu \end{bmatrix}$$

We set

$$c^T = \begin{bmatrix} \beta+\mu & \gamma+\alpha \end{bmatrix}$$

$$\begin{bmatrix} \beta+\mu & \gamma+\alpha \end{bmatrix} \begin{bmatrix} -\gamma & \alpha(1-x) \\ \beta\,(1-y) & -\mu \end{bmatrix} = \begin{bmatrix} \gamma\,\mu\,(\mathcal{R}_0-1)-(\alpha+\gamma)\,\beta\,y & \gamma\,\mu\,(\mathcal{R}_0-1)-(\beta+\mu)\,\alpha\,x \end{bmatrix}$$

We choose

$$V(x,y) = \langle X \mid c \rangle$$

where we denote $X = (x,y)^T$.

$$\dot{V} = \langle \dot{X} \mid c \rangle = \langle A(X)\,X \mid c \rangle = \langle X \mid A(X)^T\,c \rangle$$

$$\dot{V} = \left\langle \begin{bmatrix} x \\ y \end{bmatrix} \mid \begin{bmatrix} \gamma\,\mu\,(\mathcal{R}_0-1)-(\alpha+\gamma)\,\beta\,y \\ \gamma\,\mu\,(\mathcal{R}_0-1)-(\beta+\mu)\,\alpha\,x \end{bmatrix} \right\rangle$$

$$\dot{V} = \gamma\,\mu\,(x+y)\,(\mathcal{R}_0-1) - (2\,\alpha\,\beta + \alpha\,\mu + \beta\,\gamma)\,x\,y \leq 0$$

Conclusion : LaSalle

Stability of the EE

$$\mathcal{R}_0 = \frac{\alpha\,\beta}{\mu\,\gamma} > 1$$

We know that an endemic equilibirum $(\bar{x},\bar{y}) \gg 0$ exists

We use the variable change $x_{\text{new}} = x - \bar{x}$ et $y_{\text{new}} = y - \bar{y}$

$$\begin{cases} \dot{x}_{\text{new}} = & \alpha\,(y_{\text{new}} + \bar{y})\,(1 - \bar{x} - x_{\text{new}}) - \gamma\,(x_{\text{new}} + \bar{x}) \\ \\ \dot{y}_{\text{new}} = & \beta\,(1 - \bar{y} - y_{\text{new}})\,(x_{\text{new}} + \bar{x}) - \mu\,(y_{\text{new}} + \bar{y}) \end{cases}$$

To simplify we write again x for x_{new} and y for y_{new}.

Taking into account

$$\alpha\ (1 - \bar{x})\ \bar{y} - \gamma\,\bar{x} = 0$$

$$\beta\,(1 - \bar{y})\,\bar{x} - \mu\,\bar{y} = 0$$

we obtain

$$\begin{cases} \dot{x} = & -(\alpha\,\bar{y} + \gamma)\,x + \alpha\,(1 - x - \bar{x})\,y \\ \\ \dot{y} = & \beta\,(1 - y - \bar{y})\,x - (\beta\,\bar{x} + \mu)\,y \end{cases}$$

$$A(x,y) = \begin{bmatrix} -(\alpha\,\bar{y} + \gamma) & \alpha\,(1 - x - \bar{x}) \\ \\ \beta\,(1 - y - \bar{y}) & -(\beta\,\bar{x} + \mu) \end{bmatrix}$$

$$\alpha\ (1 - \bar{x})\ \bar{y} - \gamma\,\bar{x} = 0 \ \text{ in other words } \quad -(\alpha\,\bar{y} + \gamma)\,\bar{x} = -\alpha\,\bar{y}$$

$$\beta\,(1 - \bar{y})\,\bar{x} - \mu\,\bar{y} = 0 \ \text{ in other words } \quad -(\beta\,\bar{x} + \mu)\,\bar{y} = -\beta\,\bar{x}$$

$$A(x,y) = \begin{bmatrix} -\alpha\,\dfrac{\bar{y}}{\bar{x}} & \alpha - \alpha\,(x + \bar{x}) \\ \\ \beta - \beta\,(y + \bar{y}) & -\beta\,\dfrac{\bar{x}}{\bar{y}} \end{bmatrix}$$

$$\begin{bmatrix} \beta\,\bar{x} & \alpha\,\bar{y} \end{bmatrix} \begin{bmatrix} -\alpha\,\frac{\bar{y}}{\bar{x}} & \alpha - \alpha\,(x+\bar{x}) \\ \beta - \beta\,(y+\bar{y}) & -\beta\,\frac{\bar{x}}{\bar{y}} \end{bmatrix} = \begin{bmatrix} -\alpha\,\beta\,\bar{y}\,(y+\bar{y}) & -\alpha\,\beta\,\bar{x}\,(x+\bar{x}) \end{bmatrix} \ll 0$$

On

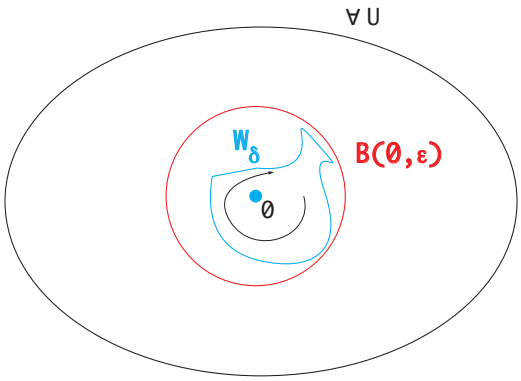
$$]\bar{x}, 1 - \bar{x}[\times]\bar{y}, 1 - \bar{y}[$$

Proof Finished with the preceding theorem.

3.3 Proofs of the Theorems

LaSalle theorem encompasses second theorem of Lyapunov. We prove Lyapunov first theorem, and after LaSalle.

Let U be an open set, $\bar{B}(x_0, \varepsilon)$ a closed ball, centered in x_0 contained in U.



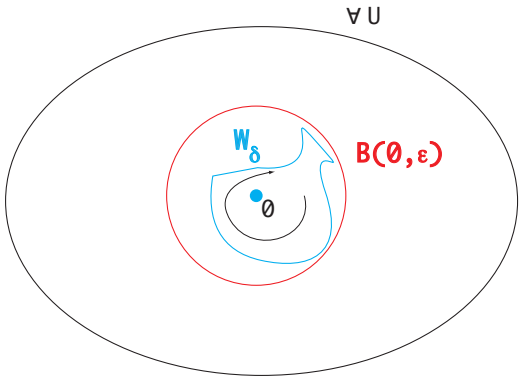
Let

$$\delta = \min_{\|x-x_0\|=\varepsilon} V(x) \ > 0$$

and

$$W_\delta = \{x \in B(x_0, \varepsilon) \mid V(x) < \delta\}$$

W_δ is an open set, $x_0 \in W_\delta \neq \emptyset$. Since V is decreasing on trajectories, a trajectory starting in W_δ cannot cross the sphere of radius ε. Remark : W_δ is positively invariant.



To prove LaSalle's invariance principle we need some concepts.

Definition 3.3.1 (invariant set) *We say that M is positively invariant for $\dot{x} = f(x)$, if for any $x_0 \in M$ we have $\Phi_t(x_0) \in M$ for any $t \geq 0$*
negatively invariant is defined similarly. A set is invariant if it is positively and negatively invariant.

Definition 3.3.2 *A point p is called an ω-limit point of the l'orbit$\gamma(x_0)$, if there exists a strictly increasing sequence of real numbers $t_1 < \ldots < t_k$ such that*

$$\lim_{k \to +\infty} x(t_k, x_0) = p$$

Theorem 3.3.1 *If the positive orbit $\gamma^+(x_0)$ is bounded then the set of ω-limits points, $\omega(\gamma)$ points is a non empty invariant compact, connected set.*

We can now prove the LaSalle's invariance principle

Let U a compact neighborhood containing x_0, on which V is a Lyapunov function.

$$m = \min_{x \in U} V(x)$$

And let

$$W = \{x \in U \mid \ V(x) \leq m\}$$

W is an invariant compact set. Any trajectory $\gamma^+(x)$ for $x \in W$ is bounded. The set of Omega-limit points of $\gamma^+(x)$ is an invariant compact set Ω_x contained in W.

The set

$$\Omega = \bigcup_{x \in W} \Omega_x$$

Is a positively invariant set.

The set Ω is an invariant set, constituted of Omega-limit point, attracting trajectories of W.

What is the value of $\dot{V}$ on Ω ?

Let $\omega \in \Omega_x$ for $x \in W$.

$$\omega = \lim_{k \to +\infty} \Phi_{t_k}(x)$$

$$V(\omega) = \lim_{k \to +\infty} V(\Phi_{t_k}(x))$$

$V(\omega)$ is a a limit value (adherence value) of $V(\Phi_t(x))$. This function is decreasing, lower bounded (by 0). Hence admits a limit c. Therefore $V(\omega) = c$, for any point of Ω_x.

$$\dot{V}(\omega) = \frac{d}{dt} V(\Phi_t(\omega)_{|t=0}$$

$$\dot{V}(\omega) = \frac{d}{dt} V(\Phi_t(\omega)_{|t=0}$$

By invariance $\Phi_t(\omega)$ is an Omega-limit point. $\Phi_t(\omega) \in \Omega_x$. Hence $V(\Phi_t(\omega)) = c$. Fonction V is constant on trajectories starting fromω.

Therefore

$$\dot{V}(\omega) = \frac{d}{dt} V(\Phi_t(\omega)_{|t=0} = 0$$

The set Ω satisfies

$$\Omega \subset \mathcal{L} = \{x \in W \mid \dot{V} = 0\}$$

This ends the proof of Theorem (3.2.2)

Now we have to prove the Theorem (3.2.3), i.e., Lasalle's Theorem for semi-definite functions.
From the proof of LaSalle's invariance principe we know that x_0 is attractive. The difficult part is to prove the stability in x_0

We restrain to $\bar{G}$ which is positively invariant and we consider the dynamical system on this compact.

$$\mathcal{L} = \{x \in \bar{G} \mid \dot{V}(x) = 0\} = \{x_0\}$$

Assume x_0 is not stable. We denote, as usual $\phi_t()$ the flow associated to the ODE. It is complete since all trajectories are bounded.

This means that we can find an $\varepsilon > 0$, a sequence of initial states x_n in G and a sequence of positive time t_n, such that

$$\|x_n - x_0\| < \frac{1}{n} \quad \text{et} \qquad \|\phi_{t_n}(x_n) - x_0\| = \varepsilon$$

We construct these element in the following way : by unstability, we know that there is ball of radius ε, such that for any ball $B(x_0, \frac{1}{n})$, there exists a x_n in this ball, such that the trajectory starting from x_n leaves the ball $B(x_0, \frac{1}{n})$. By the crossing borders theorem, there exists a time t_n such that

$$\|\phi_{t_n}(x_n) - x_0\| = \varepsilon.$$

By extracting a subsequence (we are in a compact set) $\phi_{t_n}(x_n)$, we can assume that que $\phi_{t_n}(x_n)$ converge toward a z with $\|z - x_0\| = \varepsilon$.
We claim that the sequence t_n goes to infinity, $t_n \to +\infty$. If it would not the case, sequence t_n is bounded and we can extract a subsequence t_{n_k} which converges, $t_{n_k} \to T$. But in one hand, by hypothesis

$$\lim_{k\to\infty} \phi_{t_{n_k}}(x_{n_k}) = z,$$

and in the other hand by continuity

$$\lim_{k\to\infty} \phi_{t_{n_k}}(x_{n_k}) = \phi_T(x_0) = x_0$$

Recall that x_0 is an equilibrium then for any $t > 0$ $\phi_t(x_0) = x_0$.
This is a contradiction since $\|z - x_0\| = \varepsilon$. Hence we have proved $t_n \to +\infty$ when $n \to \infty$
Since $t_n \to +\infty$, for a given $t \in \mathbb{R}$, there exists a t_n large enough such that $t_n + t > 0$. Since V is decreasing on the trajectories ($\dot{V} \leq 0$) we have

$$V(\phi_{t+t_n}(x_n) = V(\phi_t(\phi_{t_n}(x_n)) \leq V(x_n)$$

going to the limit we obtain

$$V(\phi_t(z) \leq V(x_0) \tag{3.4}$$

Now for any $s \geq 0$, again by the argument of V decreasing on trajectories we have

$$V(\phi_s(\phi_t(z)) \leq V(\phi_t(z)) \tag{3.5}$$

By attractivity of x_0, $\phi_s(\phi_t(z)) \to x_0$ when $s \to +\infty$, passing to the limit , the inequality (3.5) becomes

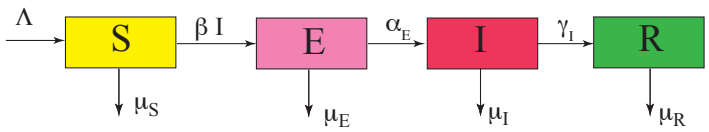
$$V(x_0) \leq V(\phi_t(z)) \tag{3.6}$$

From (3.4) and (3.6) we deduce that for any $t \in \mathbb{R}$, $V(x_0) = V(\phi_t(z))$. On the orbit of z ,

$$\gamma(z) = \{\phi_t(z) \mid t \in \mathbb{R}\}$$

V est constant.The orbit of z is invariant, hence in M, this a contradiction with avec $M = \{x_0\}$.
The equilibrium is stable and attractive $\bar{G}$.

3.4 $SEIR$ example



$$\begin{cases} \dot{S} &= \Lambda - \beta\, S\, I - \mu_S\, S \\ \dot{E} &= \beta\, S\, I - \alpha_E\, E - \mu_E\, E \\ \dot{I} &= \alpha_E\, E - \mu_I\, I - \gamma_I\, I \\ \dot{R} &= \gamma_I\, I - \mu_R\, R \end{cases} \tag{3.7}$$

R does not occur in the 3 first equations. Then we can discard the last equation

3.4.1 DFE

DFE : $\quad (\dfrac{\Lambda}{\mu_S}, 0, 0) = (S^*, 0, 0)$

$$\mathcal{R}_0 = \frac{\beta\, \alpha_E}{(\alpha_E + \mu_E)\,(\gamma_1 + \mu_I)}\,\frac{\Lambda}{\mu_S} = \frac{\beta\, \alpha_E}{(\alpha_E + \mu_E)\,(\gamma_1 + \mu_I)}\,S^*$$

Endemic equilibrium $(\bar{S}, \bar{I}, \bar{E})$ where

$$\bar{S} = \frac{(\alpha_E + \mu_E)\,(\gamma_1 + \mu_I)}{\beta\, \alpha_E} = \frac{S^*}{R_0} \qquad \bar{I} = \frac{\Lambda\,(1 - \dfrac{1}{\mathcal{R}_0})}{\beta\, \bar{S}} \qquad \bar{E} = \frac{\mu_I + \gamma_I}{\alpha_E}\,\bar{I}$$

assume (natural hypothesis)

$$\mu_S \leq \min(\mu_E, \mu_I)$$

If we denote $N = S + E + I$ we have

$$\dot{N} = \Lambda - \mu_S\, S - \mu_E\, E - (\mu_I + \gamma_I)\, I \leq \Lambda - \mu_S\, N$$

as a consequence the compact set of $\mathbb{R}^3_+$ defined by

$$\Omega = \{(S, E, I) \in \mathbb{R}^3_+ \mid S + E + I \leq \frac{\Lambda}{\mu_S} = S^*\}$$

is a positively compact invariant absorbing set
Absorbing means that any ω-limit set is in Ω.
Then we will restrict our analysis to this compact set

3.4.2 Stability of the DFE

The DFE is in Ω. Consider the Lyapunov function

$$V(S, E, I) = (\mu_I + \gamma_I)\, E + \beta\, S^*\, I$$

a simple computation gives

$$\begin{aligned}\dot{V} &= [\beta\, \alpha_E\, S^* - (\alpha_E + \mu_E)\,(\gamma_1 + \mu_I)]\, E + \beta\,(\mu_I + \gamma_I)\,(S - S^*)\, I \\ &= (\alpha_E + \mu_E)\,(\gamma_1 + \mu_I)\,[\mathcal{R}_0 - 1)]\, E + \beta\,(\mu_I + \gamma_I)\,(S - S^*)\, I \leq 0\end{aligned}$$

Since $S \leq N \leq S^*$ on Ω and $\mathcal{R}_0 \leq 1$.

What is the largest invariant set in $\mathcal{L} = \{(S, E, I) \in \Omega \mid \dot{V}(S, E, I) = 0\}$

If $\mathcal{R}_0 < 1$, necessarily $E = 0$ in $\mathcal{L}$. Invariance implies $I = 0$ hence $S = S^*$. $\mathcal{R}_0 = 1$ idem

3.4.3 Stability of the EE

Recall $\mathcal{R}_0 > 1$.

$$V(S, E, I) = (S - \bar{S}) - \bar{S}\, \ln \frac{S}{\bar{S}} + \left[(E - \bar{E}) - \bar{E}\, \ln \frac{E}{\bar{E}}\right] + \frac{\mu_E + \alpha_E}{\alpha_E}\, [(I - \bar{I}) - \bar{I}\, \ln \frac{I}{\bar{I}}\,]$$

$$\dot{V} = \Lambda - \mu_S\, S - \beta\, S\, I - \Lambda\,\frac{\bar{S}}{S} + \mu_S\,\bar{S} + \beta\,\bar{S}\, I$$
$$+\beta\, S\, I - (\alpha_E + \mu_E)\, E - \beta\, S\, I\frac{\bar{E}}{E} + (\alpha_E + \mu_E)\,\bar{E}$$
$$(\mu_E + \alpha_E)\, E - (\mu_I + \gamma_I)\,\frac{\mu_E + \alpha_E}{\alpha_E}\, I - (\mu_E + \alpha_E)\,\frac{\bar{I}}{I}\, E + (\mu_I + \gamma_I)\,\frac{\mu_E + \alpha_E}{\alpha_E}\,\bar{I}$$

$$\dot{V} = \Lambda - \mu_S\, S - \Lambda\,\frac{\bar{S}}{S} + \mu_S\,\bar{S} + \beta\,\bar{S}\, I$$
$$- \beta\, S\, I\frac{\bar{E}}{E} + (\alpha_E + \mu_E)\,\bar{E}$$
$$- (\mu_I + \gamma_I)\,\frac{\mu_E + \alpha_E}{\alpha_E}\, I - (\mu_E + \alpha_E)\,\frac{\bar{I}}{I}\, E + (\mu_I + \gamma_I)\,\frac{\mu_E + \alpha_E}{\alpha_E}\,\bar{I}$$

But recall

$$\beta\,\bar{S} = (\mu_I + \gamma_I)\,\frac{\mu_E + \alpha_E}{\alpha_E}$$

another simplification

$$\dot{V} = \Lambda - \mu_S\, S - \Lambda\,\frac{\bar{S}}{S} + \mu_S\,\bar{S}$$
$$- \beta\, S\, I\frac{\bar{E}}{E} + (\alpha_E + \mu_E)\,\bar{E}$$
$$-(\mu_E + \alpha_E)\,\frac{\bar{I}}{I}\, E + \beta\,\bar{S}\,\bar{I}$$

we write
$-\beta\, S\, I\dfrac{\bar{E}}{E} = -\beta\,\bar{S}\,\bar{I}\dfrac{\bar{E}}{E}\dfrac{S}{\bar{S}}\dfrac{I}{\bar{I}}$
$(\alpha_E + \mu_E)\,\bar{E} = \beta\,\bar{S}\,\bar{I}$ et
$-(\mu_E + \alpha_E)\,\dfrac{\bar{I}}{I}\, E = -\beta\,\bar{S}\,\bar{I}\dfrac{\bar{I}}{I}\dfrac{E}{\bar{E}}$ to get

$$\dot{V} = \beta\,\bar{S}\,\bar{I} + \mu_S\,\bar{S} - \mu_S\bar{S}\,\frac{S}{\bar{S}} - \beta\,\bar{S}\,\bar{I}\,\frac{\bar{S}}{S} - \mu_S\,\bar{S}\,\frac{\bar{S}}{S} + \mu_S\,\bar{S}$$
$$- \beta\,\bar{S}\,\bar{I}\frac{\bar{E}}{E}\,\frac{S}{\bar{S}}\,\frac{I}{\bar{I}} + \beta\,\bar{S}\,\bar{I}$$
$$- \beta\,\bar{S}\,\bar{I}\frac{\bar{I}}{I}\,\frac{E}{\bar{E}} + \beta\,\bar{S}\,\bar{I}$$

Factoring $\beta\,\bar{S}\,\bar{I}$ and $\mu_S\,\bar{S}$ we get

$$\dot{V} = \mu_S\,\bar{S}\left[2 - \frac{S}{\bar{S}} - \frac{\bar{S}}{S}\right] + \beta\,\bar{S}\,\bar{I}\,\left[3 - \frac{\bar{S}}{S} - \frac{\bar{E}}{E}\,\frac{S}{\bar{S}}\,\frac{I}{\bar{I}} - \frac{\bar{I}}{I}\,\frac{E}{\bar{E}}\right]$$

Claim : The expressions between brackets are negative definite
We have something like $2 - x - y$ with $xy = 1$ and

$3 - a - b - c$ with $abc = 1$
Function ln is concave hence

$$\frac{1}{n}\,[\ln x_1 + \cdots + \ln x_n] = \ln\,\sqrt[n]{x_1 \cdots x_n} \leq \ln[\frac{1}{n}\,(x_1 + \cdots x_n)]$$

taking exponential $n\,\sqrt[n]{x_1 \cdots x_n} - \,(x_1 + \cdots x_n) \leq 0$
if $x_1 \cdots x_n = 1$ we get

$$n - \,(x_1 + \cdots x_n) \leq 0$$

We have equality iff if all $x_i = 1$

$$\dot{V} = \mu_S\,\bar{S}\left[2 - \frac{S}{\bar{S}} - \frac{\bar{S}}{S}\right] + \beta\,\bar{S}\,\bar{I}\,\left[3 - \frac{\bar{S}}{S} - \frac{\bar{E}}{E}\,\frac{S}{\bar{S}}\,\frac{I}{\bar{I}} - \frac{\bar{I}}{I}\,\frac{E}{\bar{E}}\right]$$

This function is definite negative, this ends the proof.

blacksquare

3.5 Last example

We consider what is known a generalized Lotka-Volterra equations. This a population model. The model consist of n equations

$$\dot{x}_i = \left(r_i + \sum_{i=1}^{n} a_{ij}\,x_j\right)\,x_i \tag{3.8}$$

It is convenient to vectorialize these equations

$$\dot{x} = \mathrm{diag}(x)\,(r + A\,x)$$

Wherer $\mathrm{diag}(x)$ is the diagonal matrix whose diagonal terms are the components of x.
The term a_{ij} represent the repr?sente the incidence of species j on species i. Vector r is the birth-rate. Generaly the a_{ii} are negative, representing the intra-specific competition, as it appears in Logistic equation (also known as Pearl-Verhulst ODE).

We assume that A is a Hurwitz Metzler matrix. This is mutualism. Each species has a positive incidence on the other species. Furthermore we assume$r \gg 0$.
Then we know that there exists $\bar{x} \gg 0$ such that $Ax + r = 0$. It is sufficient to apply (5.2.2). Indeed since $-A^{-1} \geq 0$, $r \gg 0$ and $-A^{-1}$ invertible we deduce $\bar{x} = -A^{-1}\, r \gg 0$.
We have an equilibrium in the interior of the nonnnegative orthant $\bar{x}$. Now we can rewrite (3.8) as

$$\dot{x}_i = \left(\sum_{i=1}^{n} a_{ij}\, (x_j - \bar{x}_j) \right)\, x_i$$

Since A is Hurwitz, there exists $c \gg 0$ such that $A\, c \ll 0$. We now consider the Lyapunov function on the orthant

$$V(x) = \max_{i=1,\cdots,n} \frac{|x_i - \bar{x}_i|}{c_i}$$

Let i an index where $\dfrac{|x_i - \bar{x}_i|}{c_i}$ is maximum.

$$\begin{aligned}
\dot{V} &= \varepsilon_{x_i - \bar{x}_i} \frac{\dot{x}_i}{c_i} \\
&= \varepsilon_{x_i - \bar{x}_i} \frac{x_i}{c_i} \sum_{j=1}^{n} a_{ij}\, (x_j - \bar{x}_j) \\
&= \frac{x_i}{c_i} \left[a_{ii}\, |x_i - \bar{x}_i| + \sum_{j \neq i} \varepsilon_{x_i - \bar{x}_i}\, a_{ij}\, (x_j - \bar{x}_j) \right] \\
&\leq \frac{x_i}{c_i} \left[a_{ii}\, |x_i - \bar{x}_i| + \sum_{j \neq i} a_{ij}\, |x_j - \bar{x}_j| \right] \\
&\leq \frac{x_i}{c_i} \left[a_{ii}\, |x_i - \bar{x}_i| + \sum_{j \neq i} a_{ij}\, \frac{c_j}{c_i} |x_i - \bar{x}_i| \right] \\
&= \frac{x_i}{c_i^2}\, |x_i - \bar{x}_i| \left[a_{ii}\, c_i + \sum_{j \neq i} a_{ij}\, c_j \right] \\
&= \frac{x_i}{c_i}\, V(x)\, (A\, c)_i \leq 0
\end{aligned}$$

Function $\dot{V}$ is negative definite. this end s the proof of the GAS of the equilibrium..

3.6 Reduction of systems and Vidyasagar's Theorem

We consider a triangular system, more precisely this is a system on $\mathbb{R}^n \times \mathbb{R}^m$ which can be written in the following form

$$\begin{cases} \dot{x}_1 &= f_1(x_1) \\ \dot{x}_2 &= f_2(x_1, x_2) \end{cases} \tag{3.9}$$

where f_1 is an application from $\mathbb{R}^n$ into $\mathbb{R}^n$ and f_2 from $\mathbb{R}^n \times \mathbb{R}^m$ into $\mathbb{R}^m$. We will assume that the conditions for existence and uniqueness are satisfied(for example f_1 and f_2 of class $\mathcal{C}^1$).
The trajectories have the same projection on $\mathbb{R}^n \times \{0\}$, these are the ones of system $\dot{x}_1 = f_1(x_1)$ sur $\mathbb{R}^n$.
It is clear why they are called triangular. Actually the Jacobian is lower block triangular. Theses systems are also called hierarchical. We will give a result obtain by Vidyasagar [97]. We give the autonomous version which is more simple. However this theorem is also valid, with some extra-hypothesis for non-autonomous systems.

Theorem 3.6.1 *[Vidyasagar]*
We consider a $\mathcal{C}^1$ system

$$\begin{cases} \dot{x}_1 &= f_1(x_1) \\ \dot{x}_2 &= f_2(x_1, x_2) \end{cases} \tag{3.10}$$

such that the origin of $\mathbb{R}^n$ is globally asymptotically stable (GAS) for the isolated system $\dot{x}_1 = f_1(x_1)$ on $\mathbb{R}^n$ and such that the origin of $\mathbb{R}^m$ is GAS pour $\dot{x}_2 = f_2(0, x_2)$.
Then the origin is asymptotically stable.
If all the positive trajectories are bounded then the origin is GAS on $\mathbb{R}^n \times \mathbb{R}^m$.

This theorem is very convenient because it allows to "reduce" the system. For example if the stability of the isolated system is obvious, it remains to study the reduced system $\dot{x}_2 = f_2(0, x_2)$.

Remark 3.6.1 *If all the positive trajectories are not bounded, only the local asymptotic stability can be obtained. The following example in $\mathbb{R}^2$, from Seibert and Suarez [81], is a counter-example.*

$$\begin{cases} \dot{x} &= -x \\ \dot{y} &= y\,(x^2\,y^2 - 1) \end{cases} \tag{3.11}$$

It is clear that 0 is GAS for the first system $\dot{y} = -y$. Since we have symmetries, we consider the system in the nonnegative orthant. Let the functions $H_K(x, y) = x\,y - k$

$$\dot{H}_k = \left\langle \nabla H_k \;\middle|\; \begin{bmatrix} -x \\ y\,(x\,y - 1) \end{bmatrix} \right\rangle = x\,y\,(x^2\,y^2 - 2)$$

Hyperbolas $xy - \sqrt{2} = 0$ are invariant. These hyperbolas $xy - a$ where $a > \sqrt{2}$ are boundaries where the vector field points toward the increasing xy. The origin is not GAS.

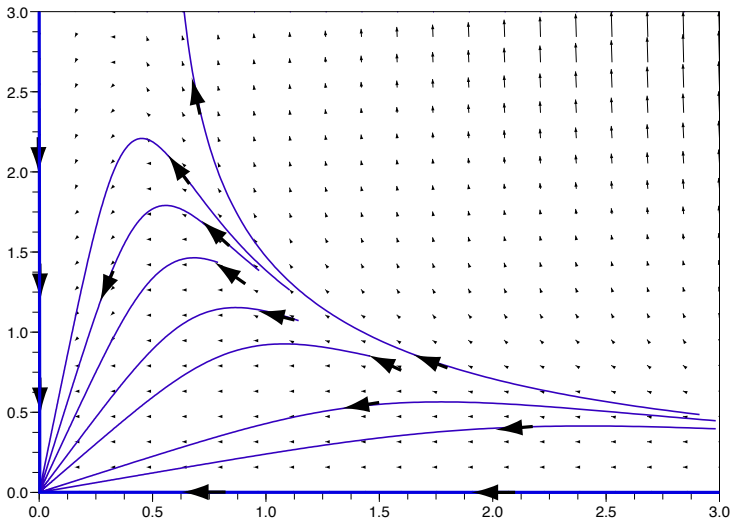


Figure 3.6: Local Asymptotic Stability only

Proof

We prove stability. Let

$B(0, \varepsilon) = \{(x_1, x_2) \mid \|x_1\| \leq \varepsilon, \|x_2\| \leq \varepsilon\}$ a neighborhood of the origin.

Since the equilibria of the isolated systems are GAS, since these systems are $\mathcal{C}^1$ the converse of Lyapunov Theorems can be applied. Then it exist $\mathcal{C}^1$ positive definite functions $V_1(x_1)$ and $V_2(x_2)$ such that

$$\dot{V}_1 = \langle \nabla V_1(x_1) | f_1(x_1) \rangle \leq 0$$

$$\dot{V}_2 = \langle \nabla V_2(x_2) | f_2(0, x_2) \rangle \leq 0$$

These functions $\dot{V}_1$ et $\dot{V}_2$ are negative definite on $B(0, \varepsilon)$ for ε small enough.
Since f_1 et V_1 are $\mathcal{C}^1$ let

$$L = \max_{(x_1,x_2)\in B(0,\varepsilon)} \frac{\partial f_1}{\partial x_1}(x_1, x_2)$$

and

$$M = \max_{(x_1,x_2)\in B(0,\varepsilon)} \nabla V_2(x_2)$$

Since V_2 is a Lyapunov function, we can choose $\delta_1 < \frac{\varepsilon}{2}$ small enough such that

$$\max_{\|x_2\|\leq\delta_1} V_2(x_2) \; < \; \min_{\frac{\varepsilon}{2}\leq\|x_2\|\leq\varepsilon} V_2(x_2)$$

We have

$$\dot{V}_2(x_2) = \langle \nabla V_2(x_2) | f_2(x_1, x_2) \rangle = \langle \nabla V_2(x_2) | f_2(0, x_2) \rangle + \langle \nabla V_2(x_2) | f_2(x_1, x_2) - f_2(0, x_2) \rangle$$

With the relation

$$f_2(x_1, x_2) - f_2(0, x_2) = \int_0^1 \frac{\partial f_2}{\partial x_1}(t\, x_1, x_2)\, x_1\, dt$$

which gives in $B(0, \varepsilon)$

$$\|f_2(x_1, x_2) - f_2(0, x_2)\| \leq L\, \|x_1\|$$

and

$$\dot{V}_2(x_2) \leq \langle \nabla V_2(x_2) | f_2(0, x_2) \rangle + L\, M\, \|x_1\| \tag{3.12}$$

Function $\langle \nabla V_2(x_2) | f_2(0, x_2) \rangle$ is negative definite, therefore if we define φ by

$$\varphi(c) = \min_{c\leq\|x_2\|\leq\varepsilon} -\langle \nabla V_2(x_2) | f_2(0, x_2) \rangle$$

The function φ, defined in $\mathbb{R}$, is continuously increasing , tends to 0 when c tends to 0 and satisfies $\varphi(c) > 0$ pour tout $c > 0$.
Since $\dot{x}_1 = f_1(x_1)$ is AS, we can choose $\delta_2 \leq \delta_1$ such that, if the initial condition satisfies $\|x_1(0)\| \leq \delta_2$, we will have for any $t \geq 0$, the inequality $\|x_1(t)\| \leq \frac{\varphi(\delta_1)}{L\, M}$.
If we have $\|x_1\| \leq \delta_2$ and $\|x_2\| \geq \delta_1$, with the inequality (3.12), we deduce

$$\langle \nabla V_2(x_2) | f_2(0, x_2) \rangle + L\, M\, \|x_1\| < 0 \tag{3.13}$$

Let now δ_3 be such that $0 < \delta_3 < \delta_2$ and such that

$$\max_{\|x_1\| \leq \delta_3} V_1(x_1) \ < \ \min_{\delta_2 \leq \|x_1\| \leq \varepsilon} V_1(x_1)$$

Let $\mathcal{U}$ the open set defined by

$$\mathcal{U} = \{(x_1, x_2) |\ \|x_1\| \leq \delta_3; \|x_2\| \leq \delta_3\}$$

If $x_1(0) \leq \delta_3$, since V_1 is decreasing, the preceding inequaltiy shows that $\|x_1(t)\| \leq \delta_2$. In other words no trajectory can reach the sphere of radius δ_2 in $\mathbb{R}^n$).
Let now $\|x_2(0)\| \leq \delta_3$. Since

$$\max_{\|x_2\| \leq \delta_3} V_2(x_2) \ \leq \ \max_{\|x_2\| \leq \delta_1} V_2(x_2) \ < \ \min_{\frac{\varepsilon}{2} \leq \|x_2\| \leq \varepsilon} V_2(x_1)$$

the trajectory starting from $(x_1(0), x_2(0))$ as long it verifies $\|x_2(t)\| \leq \delta_1$

$$V_2(x_2(t)) \leq \min_{\delta_1 \leq \|x_2\| \leq \varepsilon} V_2(x_1)$$

We have seen above that we have $\|x_1(t)\| \leq \delta_2$. This implies, as long as $\|x_2(t)\| \geq \delta_1$, from the inequality (3.13), and inequality $\dot{V}_2(x_2) \leq 0$.
Since V_2 is non-increasing on the trajectories in the ring defined by $\|x_1\| \leq \delta_2$, $\delta_1 \leq \|x_2\| \leq \frac{\varepsilon}{2}$, we conclude that any trajectory cannot reach the sphere of radius $\frac{\varepsilon}{2}$ in $\mathbb{R}^m$. We proved $\|x_1(t)\| \leq \delta_2 < \varepsilon$ and $\|x_2(t)\| \leq \frac{\varepsilon}{2}$. This ends the stability proof.
We will show the local attractivity by LaSalle's invariance principle.
Since the origin is stable, there exists a compact neighborhood of the origin $\mathcal{U}$, positively invariant. we will restrict ourselves to this neighborhood $\mathcal{U}$.

Let V_1 a Lyapunov-LaSalle functionBy hypothesis

$$\dot{V}_1 = \langle \nabla V_1(x_1) | f_1(x_1) \rangle \leq 0$$

Let $E = \{(x_1, x_2) \in \mathcal{U} \mid \dot{V}_1(x_1) = 0\}$ and the greatest invariant set contained in E. This is clearly $(\{0\} \times \mathbb{R}^m) \cap \mathcal{U}$. By hypothesis $\dot{x}_2 = f_2(0, x_2)$ is GAS on $\{0\} \times \mathbb{R}^m$. We claim that this implies that any negative trajectory from $\mathcal{U} \setminus \{0\}$ goes out of $\mathcal{U}$.
Indeed, if it is not the case, we will have a complete trajectory γ in $\mathcal{U}$. The set of α-limit points of γ is invariant. By asymptotic stability and invariance, this set contain the origin. This means that trajectory starts as close as we want of the origin to going back to this origin. The closure of this trajectory is compact, which contradicts the stability. This proves our claim.
This means that the the greatest invariant set contained in E is reduced to $\{0\}$. This proves the attractiveness of the origin in $\mathcal{U}$.

If a trajectory is relatively compact, then the ω-limit points are in $\{0\} \times \mathbb{R}^m$. Indeed for $t_n \to \infty$ we have $x_1(t_n) \to 0$. If all trajectories are relatively compact the ω-limits points are in $\{0\} \times \mathbb{R}^m$. By asymptotic stability on $\{0\} \times \mathbb{R}^m$ the origin is an ω-limit point. Any trajectory approaches as close as we wont to the origin. By stability it is trapped in $\mathcal{U}$ above. The tends to the origin.

■

Chapter 4

The concept of basic reproduction ratio $\mathcal{R}_0$

4.1 Introduction

This chapter consists essentially of two parts. A mathematical part, which uses mathematical theorems and demonstrations. Finally, examples to understand the notion.

We denote by $\mathcal{R}_0$ the basic reproduction rate. This concept is now unanimously recognized as a key concept in epidemiology. It is defined "heuristically" as the average number of new cases of infection, caused by an average infected individual (during the infective period), in a population entirely composed of susceptible.

For about twenty years, $\mathcal{R}_0$ is part of the majority of research articles using mathematical modeling.

Originally, this concept comes from demography and ecology. This is the average number of girls (females) born of a female (female) during her life. The use of $\mathcal{R}_0$ is relatively recent in epidemiology.

The first to have introduced this concept in 1886 is undoubtedly Richard B?ckh, the director of the statistics office of Berlin. Using a life table for women from 1879, it summed the probability of survival products by the birth rate of girls. He concludes that, on average, 2,172 girls will be born to a woman. He corrects this figure using the sex ratio and comes up with, what we would call $\mathcal{R}_0$, an estimate of 1.06.

It is Dublin and Lotka (1925) and Kuczynski (1928), which introduce, in the demographic context, the notion and calculation of $\mathcal{R}_0$. In a summary in 1939, in French, of his contribution Lotka writes

> The net reproductivity, $\mathcal{R}_0$, introduced by Boeckh, has more merit, since it gives a measure essentially independent of the age distribution of the population.

If $\mathcal{F}(a)$ is the probability for a woman to survive at a, if $\beta(a)$ is the birth rate of girls, at age a for a woman then

$$\mathcal{R}_0 = \int_0^{+\infty} \mathcal{F}(a)\,\beta(a)\,da.$$

This is the approximation calculation done by B?ckh. This is the mean number of girls that woman will have during all her life.
One wonders where the 0 index comes from. We can define the order time n for the function $\mathcal{F}(a)\,\beta(a)$

$$\mathcal{R}_n = \int_0^{\infty} a^n\,\mathcal{F}(a)\,\beta(a)\,da$$

The basic reproduction rate is the 0 moment of order.
The concept of threshold has been used by Ross in his elementary mosquito theorem. Neither Ross nor Kermack and McKendrick have attached a name or symbol to their threshold concept. From 1911, Ross, Ross and Hudson developed a theory of epidemics between 1916 and 1917. Ross calls it a priori pathometry or theory of "happenings". It should be noted that Lotka was very interested in Ross papers and solves several of the problems posed by Ross in 1919 in "a contribution to quantitative epidemiology". He also devotes 120 pages "contribution to malaria epidemiology" to the model of Ross [69]. In his article on the history of $\mathcal{R}_0$ [39], Heesterbeek regrets that Lotka, who introduced $\mathcal{R}_0$ demographer and biomathematician, missed out on this concept in epidemiology.
McKendrick was also a military doctor. He served under Ross in the 1901 Sierra Leone eradication campaign. It was Ross who encouraged McKendrick to learn mathematics and apply it to medical problems. In his correspondence with McKendrick, Ross makes clear his desire to establish the general law of epidemics. In 1911 he wrote a letter to McKendrick

> We shall end by establishing a new science. But first let you and me unlock the door and then anybody can go in who likes.

The following basic paper is still that of Kermack and McKendrick [57]. They establish in the continuity of Ross a threshold of critical density N_c for the population in the form

$$\frac{1}{N_c} = \int_0^{\infty} \phi_t\,B_t\,dt$$

The next step is given by McDonald in 1952. George McDonald is the first professor of Tropical Medicine and Hygiene and director of the Ross Institute in 1947. He has had a considerable influence on the use of modeling in malariology. It introduces the term basic reproduction rate.
In 1975 Dietz and Hethcote rediscover the $\mathcal{R}_0$ concept for direct transmission diseases. Dahlem's 1982 conference, initiated by May and Anderson, serves to promote the concept. We must finally wait for Diekmann et al. to give a precise mathematical definition It is defined an operator, called the "next generation operator". The largest eigenvalue defined $\mathcal{R}_0$. This operator is for a generation.
The classical and non-mathematical definition of $\mathcal{R}_0$ is, as given in the introduction, the average number of secondary cases of infectious disease, generated by a typical individual in a population consisting entirely of susceptible individuals throughout their entire period of infectivity.
This is a heuristic definition. Diekmann and Heesterbeek gave a mathematical definition. In these notes we will treat the case of compartmental deterministic models in finite dimension.

4.2 The structure of compartmental epidemiological models

We will make an essential assumption: there is no immigration of infected individuals. Indeed, if this were the case, there would be no balance without disease.

We will consider that the population is divided into n compartments. The number of individuals in the i compartment is given by x_i (or x_i can be a prevalence , i.e. the percentage of individuals contained in the i compartment relative to the total population, or a density)
It is assumed that the compartments are marked in such a way that the first p consist of "uninfected" individuals, more specifically non-carriers of the germ (virus, protozoan, parasite, dots). In fact all those who will not evolve to a compartment of infectious individuals. In these compartments may have susceptible, vaccinated, quarantined individuals in such a way that they can transmit neither horizontally nor vertically.
The essential concept is that these compartments will never be able to give transmitters of their own.
The rest of the compartments are infected. For example infectious, latent, asymptomatic carrier.
we denote by

$$x = (x_1, x_2, \cdots, x_n)^T$$

the state of the system

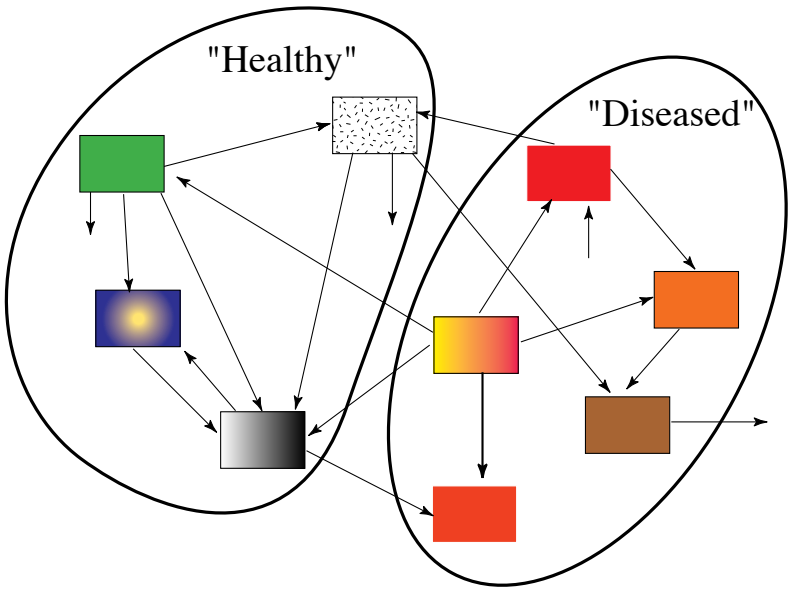


Figure 4.1: infected and non infected

We can now describe the dynamics of the system
In other words we will rewrite the ODE $\dot{x} = f(x)$. Let the compartment x_i.

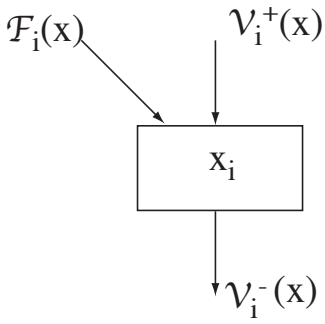


Figure 4.2: Le bilan

We will now describe the dynamics of this infectious disease.

In other words, we will write the differential equation $\dot{x} = f(x)$. Let the compartment x_i.
We consider the balance of what comes in and what comes out in each compartment:

1. We denote by $\mathcal{F}_i(x)$ the speed of appearance of new infected, in the compartment i. They are new infected, obtained by transmission of any kind. Horizontal, i.e., from individual to individual or vertical from mother to child.

2. We denote by $\mathcal{V}_i^+(x)$ is from other compartments by any other cause (moving, aging, healing etc ...)

3. We denote by $\mathcal{V}_i^-(x)$ the speed of what leaves the i compartment. For example by mortality, change of epidemiological status, movement etc ...

We finally have

$$\dot{x}_i = \mathcal{F}_i(x) + \mathcal{V}_i^+(x) - \mathcal{V}_i^-(x)$$

X_s are states without disease, i.e.

$$X_s = \{x \mid x_{p+1} = \cdots = x_n = 0\}.$$

The nature of the epidemiological characteristics implies the following properties for the introduced functions:

H1 $x \geq 0$ and $\mathcal{F}(x) \geq 0$, $\mathcal{V}_i^+(x) \geq 0$, $\mathcal{V}_i^-(x) \geq 0$

Indeed this is flows of materials.

H2 If $x_i = 0$ then $\mathcal{V}_i^-(x) = 0$

If there is nothing in a compartment, nothing can come out of it, it is the essential property of a compartmental model.

H3 If $i \leq p$ then $\mathcal{F}_i(x) = 0$

Compartments with an index of less than p are "uninfected". By definition, it can not appear in these compartments "infected".

H4 If $x \in X_s$ then $\mathcal{F}_i(x) = 0$ and for $i \geq p$ we have $\mathcal{V}_i^+(x) = 0$

Recall that we have assumed that there is no immigration of infectives. Since we are in a state in X_s this means that we are in the state with no infected anywhere. $\mathcal{V}_i^+(x)$ is infectious coing from other compartment.

Nothing can move from a uninfected compartment in i, since it is an infected compartment. And all the infected compartment are empty.

If there are no carriers of germs, in the population, no new "infected" can appears. This is Lavoisier's principle. There is no spontaneous generation.

We will consider a point of equilibrium without disease, which is also a point of equilibrium of the system. For example, in the case of demographic dynamics, this means that the population is not moving. In fact, every x_i^* is fixed and is zero for $i > p$. In other words, a "disease-free" equilibrium, $x^* \in X_s$ is such that $f(x^*) = 0$ Such an equilibrium is called a disease-free equilibrium (DFE).
We denote , for any state, by $\mathbf{x}_1 = (x_1, \cdots, x_p)^T$ the "non infected" components $\mathbf{x}_2 = (x_{p+1}, \cdots, x_n)^T$) the "infected" one.
The system $\dot{x} = f(x)$ can be rewritten

$$\begin{cases} \dot{\mathbf{x}}_1 &= f_1(\mathbf{x}_1, \mathbf{x}_2) \\ \dot{\mathbf{x}}_2 &= f_2(\mathbf{x}_1, \mathbf{x}_2) \end{cases} \tag{4.1}$$

By definition $x^* = (\mathbf{x}_1^*, 0)$ where

$$f_1(\mathbf{x}_1^*, 0) = 0$$

and for any x_1 we have

$$f_2(x_1, 0) = 0$$

Indeed if $(x_1, 0)$ is a state without disease state, any new infected can appear.
If we assume that f is $\mathcal{C}^1$ then there are matrices $A_{11}(x)$, $A_{12}(x)$, and $A_{22}(x)$ such that

$$f_1(\mathbf{x}_1, \mathbf{x}_2) = A_{11}(x).(\mathbf{x}_1 - \mathbf{x}_1^*) + A_{12}(x).\mathbf{x}_2$$

and

$$f_2(\mathbf{x}_1, \mathbf{x}_2) = A_{22}(x).\mathbf{x}_2$$

These two results result from the proposition 1.2.1 applied to the functions $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$. According to this proposition we know that there exists a matrix $A_1(x)$ of size $p \times n$ such that

$$f_1(x) = f_1(x_1, x_2) = A_1(x) \begin{bmatrix} x_1 - x_1^* \\ x_2 \end{bmatrix}$$

We break A_1 into a block A_{11} of size $p \times p$ and a block A_{12} of size $p \times n - p$

$$A_1 = \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

This gives the first statement.
The second relationship comes from the proposition applied to the function $f_2(x)$, considered as a function of x_2, which vanishes in 0.
The system is rewritten

$$\begin{cases} \dot{\mathbf{x}}_1 &= A_{11}(x).(\mathbf{x}_1 - \mathbf{x}_1^*) + A_{12}(x).\mathbf{x}_2 \\ \dot{\mathbf{x}}_2 &= A_{22}(x).\mathbf{x}_2 \end{cases} \tag{4.2}$$

The Jacobian at the point of equilibrium $(\mathbf{x}_1^*, 0)$ is written

$$\text{Jac}(x^*) = \begin{bmatrix} A_{11}(x^*) & A_{12}(x^*) \\ 0 & A_{22}(x^*) \end{bmatrix}$$

noindent We will make an additional assumption We suppose that

H5 The DFE is $\mathbf{x}^* = (\mathbf{x}_1^*, 0)$. If $\mathcal{F}(x)$ is set to zero, the matrix $Df(\mathbf{x}^*)$ is Hurwitz.

It simply means that when there is no disease, the population admits a locally asymptotically stable equilibrium, the DFE.
We have $J(x^*) = D\mathcal{F}(x^*) + D\mathcal{V}^+(x^*) + D\mathcal{V}^-(x^*)$. Moreover, since the $\mathcal{F}_i$ components of the function $\mathcal{F}$ are identically zero for $i \leq p$ we have

$$D\mathcal{F}(x^*) = \begin{bmatrix} 0 & 0 \\ 0 & F \end{bmatrix}$$

and

$$D(\mathcal{V}^+ - \mathcal{V}^-)(x^*) = \begin{bmatrix} J_3 & J_4 \\ 0 & V \end{bmatrix}$$

Theorem 4.2.1 *$F \geq 0$ and V is an asymptotically stable Metzler matrix*

Proof
We denote by e_i the i-th vector of the canonical basis of $\mathbb{R}^n$. For $i > p$ and $j > p$ since $\mathcal{F}_i(x^*) = 0$

$$\frac{\partial \mathcal{F}_i}{\partial x_j}(x^*) = \lim_{h \to 0^+} \frac{\mathcal{F}_i(x^* + h\, e_j)}{h} \geq 0$$

For indices satisfying $i > p$ and $j > p$ and $i \neq j$ we have $\mathcal{V}_i^-(x^* + h\, e_j) = 0$. Indeed $x^* + h\, e_j$ is a state wher we add to x^*, h element from the infected compartments j, with $j \neq i$. So there is nothing in the compartment i when system is in state $x^* + h\, e_j$, therefore nothing can go out. Hence $\mathcal{V}_i(x^* + h\, e_j) = \mathcal{V}_i^+(x^* + h\, e_j)$. Now since $i > p$ we have $\mathbf{x}_i^* = 0$ ($\mathbf{x}^*$ is a DFE)

We deduce

$$\frac{\partial \mathcal{V}_i}{\partial x_j}(x^*) = \lim_{h \to 0^+} \frac{\mathcal{V}_i^+(x^* + h\, e_j)}{h} \geq 0$$

This proves that V is Metzler. By **H5** matrices J_3 and V are asymptotically stable . Matrix V is an asymptotically stable matrix..

■

4.2.1 Definition of $\mathcal{R}_0$

Definition 4.2.1 (spectral radius) *:*
We call the spectral radius of a matrix A, the maximum value of the module of the eigenvalues ?? of A.

$$\rho(A) = \max_{\lambda \in Sp(A)} |\, \lambda \,|$$

Definition 4.2.2 ($\mathcal{R}_0$)

$$\mathcal{R}_0 = \rho(-F\, V^{-1})$$

First like $F \geq 0$ and V is an asymptotically stable Metzler, then $-V^{-1} \geq 0$. This is demonstrated in the theorem 5.2.2. This proves that $-F\, V^{-1}$ is a positive matrix. According to the classic Perron-Frobenius theorem, the spectral radius is an eigenvalue of this matrix.
This definition is purely mathematical.

Remark 4.2.1 *The definition of "next generation matrix" given here differs from a $-$ sign compared to that of van den Driessche [94]. We use Metzler matrices, which appear naturally in compartmental systems, whereas in [94] the M -matrices are used. Which leads van den Driessche to note $\mathcal{V}_i^+$ what comes in, $\mathcal{V}_i^-$ what comes out and to note*

$$\dot{x}_i = \mathcal{F}_i - \mathcal{V}_i$$

with $\mathcal{V}_i = \mathcal{V}_i^- - \mathcal{V}_i^+$!!!
This is absolutely unnatural, but it is to bring up the opposites of the stable Metzler matrices, the M -matrices

4.2.2 Biological interpretation of $\mathcal{R}_0$

We will now give the biological interpretation of the definition of $\mathcal{R}_0$.
A small number of infectious individuals are introduced into a susceptible population. We are therefore at equilibrium at the DFE.
To determine the fate of a small number of infected individuals, we consider the dynamics of the system, with reinfection suppressed, since we are interested in the evolution of the introduced infectious individuals. As we want their immediate future, we consider the system approached by its linearization at equilibrium. If one is close to equilibrium the behavior of the system is approximated by the linearized system.
Since $Df = D\mathcal{F} + D\mathcal{V}^+ - D\mathcal{V}^-$ the system becomes

$$\dot{x} = (D\mathcal{V}^+ - D\mathcal{V}^-).x = \begin{bmatrix} J_3 & J_4 \\ 0 & V \end{bmatrix}.x$$

If $(0, \mathbf{x}_2^0)$ is a small number of infected individuals at time t we have, by integrating the linear system, to the time $e^{tV}.\mathbf{x}_2^0$ infectious people. This represents the state of the infected in the infected compartments.
In the end we will have obtained

$$\int_0^\infty (0, e^{tV}.\mathbf{x}_2^0)\, dt = (0, -V^{-1}.\mathbf{x}_2^0)$$

This set of infected will generate new cases by transmission. The number of new cases will be

$$-F\,V^{-1}.\mathbf{x}_2^0$$

We will interpret the components of

$$-F\,V^{-1}$$

If we consider an infected in the j compartment, then the (i, j) entry of $-V^{-1}$ is the average time that this individual will stay in the i compartment during its "infective period". The (k, i) entry of F is the speed with which an individual in the i compartment produces new infections in the k compartment. Therefore the (k, j) entry of $-FV^{-1}$ is the average number of new infections in the k compartment produced by an infected individual in the j compartment.
If K is the next generation matrix, we have just seen that $K_{i,j}$ is the average number of infected individuals of type i produced by an infected individual of type j. The K matrix is a positive square matrix of the size dimension of the type of infected.

The matrix $-FV^{-1}$ is called the "next generation matrix". Approximately $-FV^{-1}\,x_0$, vectorially expressed, the "number" of new secondary cases. We are led to consider, at generation n, the quantity $(-FV^{-1})^n\,x_0$. In other words $M^n.x_0$ where M is a positive operator. It is a discreet positive system. The importance of dominant modes in these positive systems is well known. The term mode is the term of the engineers or physicists to designate the eigenvalues. Dominant mode is simply the spectral radius. By the Perron-Frobenius theorem ref Perron, it is a proper value. Hence the definition of $mathcalR_0$. The system is stable and converges to 0 if $mathcalR_0 < 1$. The system is unstable and the state of the system tends to infinity, and is aligned with the eigenvector corresponding to the largest eigenvalue. This is an intuitive result. But we have a lot more. This is the subject of the next part.

4.3 $\mathcal{R}_0$ is a threshold

For a dynamic system, we call threshold at the point of equilibrium a function of the parameters of the system $\mathcal{T}$ such that if $\mathcal{T} < 1$ then the system is locally asymptotically stable and unstable if $\mathcal{T} > 1$.

Theorem 4.3.1 *The epidemiological system is asymptotically stable to the DFE if* $\mathcal{R}_0 < 1$ *and unstable if* $\mathcal{R}_0 > 1$.

Proof
We apply the Poincar?-Lyapunov theorem of linearization Just look at the Jacobian in x^*:

Proof
We apply Poincar?-Lyapunov linearization theorem. It is sufficient to consider the Jacobian at x^* :

$$\mathrm{Jac}(x^*) = \begin{bmatrix} A_{11}(x^*) & A_{12}(x^*) \\ 0 & A_{22}(x^*) \end{bmatrix} = \begin{bmatrix} J_3 & J_4 \\ 0 & F+V \end{bmatrix}$$

By the hypothesis **H5** $A_{11}(x^*) = J_3$ is Hurwitz. Therefore it is sufficient to prove that $F+V$ is Hurwitz. But $F+V$ is a regular decomposition (see definition 4.3.1) of $A_{22}(x^*)$.

A regular decomposition of a matrix is the decomposition of this matrix in the sum of a nonnegative matrix (here F) with a Hurwitz Metzler matrix (here V). Varga's theorem (See Theorem 4.3.2) gives the equivalence :

$s(F+V) < 0$ is equivalent to $\rho(-FV^{-1}) < 1$.

By continuity $s(F+V) \leq 0$ is equivalent to $\rho(-FV^{-1}) \leq 1$. This implies

$s(F+V) > 0$ is equivalent to $\rho(-FV^{-1}) > 1$.

Conclusion is obtain by Poincar?-Lyapunov theorem. ■

4.3.1 Varga's Theorem

We will prove a theorem of Varga [95, 96] which is closely related to $\mathcal{R}_0$.

Definition 4.3.1 (regular decomposition)
Let A a Metzler matrix.A regular decomposition of A is any decomposition of A

$$A = F + V$$

where $F \geq 0$ and V is Hurwitz Metzler matrix.

Theorem 4.3.2
For any regular decomposition of a Metzler matrix A, the following assertions are equivalent

- *A is Hurwitz*
- *$\rho(-F\,V^{-1}) < 1$*

Remark 4.3.1
Any regular decomposition gives a threshold

Proof
Suppose that A is Hurwitz, then we claim $-A^{-1} \geq 0$ (see below)
Matrices $V = A - F$ and A being invertible (note that a Hurwitz matrix is non-singular) we have using $A - F = (I - F\,A^{-1})\,A$

$$-F\,V^{-1} = -F\,(A-F)^{-1} = -F\,A^{-1}\,(I - F\,A^{-1})^{-1}$$

Let $G = -F\,A^{-1}$. This is a nonnegative matrix. To obtain its spectral radius, from Perron-Frobenius, it s sufficient to consider only nonnegative vectors. Let $v > 0$ an eigenvector of G associated to the eigenvalue $\lambda \geq 0$, such that $G\,v = \lambda\,v$. We have

$$-F\,V^{-1}\,v = G(I+G)^{-1}\,v = \frac{\lambda}{1+\lambda}\,v.$$

Matrix $-F\,V^{-1}$ is nonnegative. Reciprocally let $\mu \geq 0$ an eigenvalue of $-F\,V^{-1}$ associated to an eigenvector $v > 0$. Then $G(I+G)^{-1}\,v = \mu\,v$. Since G and $(I+G)^{-1}$ commute, we deduce $G\,v = \mu\,(I+G)\,v$ or $(1-\mu)\,G\,v = \mu\,v$. This implies that necessarily $\mu \neq 1$ and v is an eigenvector of G associated to the eigenvector to $\dfrac{\mu}{1-\mu}$.

The function from $\mathbb{R}^+$ into $[0,1[$, defined by par $x \mapsto \dfrac{x}{1+x}$ is a bijection from the eigenvalues of $G = -F\,A^{-1}$ onto the eigenvalues of $-F\,V^{-1}$. This a monotonous function. Therefore

$$\rho(-F\,V^{-1}) = \frac{\rho(G)}{1+\rho(G)} \;<\; 1$$

Reciprocally suppose that $\rho(-F\,V^{-1}) < 1$. Then $-I - F\,V^{-1}$ is invertible, and is a Metzler matrix. Since $\rho(-F\,V^{-1}) < 1$ we have $s(-I - F\,V^{-1}) < 1$. This is a Hurwitz Metzler. Then the opposite of its inverse is nonnegative therefore

$$-A^{-1} = (-I - F\,V^{-1})^{-1}\,V^{-1} \geq 0$$

This proves that A est Hurwitz. This ends the proof.

■

Lemma 4.3.1
Let A a Metzler matrix, the following assertions are equivalent

- *A is Hurwitz*
- *$-A^{-1} \geq 0$.*

Proof
Assume A is Hurwitz then

$$-A^{-1} = \int_0^{+\infty} e^{tA}\,dt = \left[A^{-1}e^{tA}\right]_0^{+\infty}$$

Let e_i the canonical basis of $\mathbb{R}^n$

$$(-A^{-1})_{i,j} = \langle {}^{-1}_{A}\,e_j \mid e_i\rangle = \int_0^{+\infty} \langle e^{tA}\,e_j \mid e_i\rangle\,dt \geq 0$$

We have used that e^{tA} let invariant the nonnegative orthant (use barrier Theorem 3.1.1)
Reciprocally assume $-A^{-1} \geq 0$. Let $c \gg 0$ then $v = -A^{-1}\,c \gg 0$ this gives $A\,v = -c \ll 0$.
Consider the ODE $\dot{x} = A^T\,x$.

Let now $V(x) = \langle v \mid x\rangle$. This is a Lyapunov function on the nonnegative orthant. For $x \geq 0$

$$\dot{V}(x) = \langle v \mid A^T\,x\rangle = \langle A\,v \mid x\rangle = -\langle c \mid x\rangle \leq 0$$

and $\dot{V}$ is definite negative on $\mathbb{R}^n_+$. Hence the ODE is asymptotically stable on the nonnegative orthant.

Now let $x \in \mathbb{R}^n$ and $x = x^+ - x^-$. We have $e^{tA} = e^{tA}\,x^+ - e^{tA}\,x^- \to 0$ when $t \to +\infty$. This proves that A is Hurwitz. (The origin is attractive, and A is linear) ■

We have more, by continuity of the functions s and ρ we obtain

$$s(F+V) \leq 0 \Longleftrightarrow \rho(-FV^{-1}) \leq 1$$

By contraposition we obtain

$$s(F+V) > 0 \Longleftrightarrow \rho(-FV^{-1}) > 1.$$

4.4 Examples

We give examples to illustrate the computation of $\mathcal{R}_0$. We will consider classical examples.

Ross model

We consider the model with prevalences. in other words x represent the % of infected humans and y the % of infected mosquitoes.

$$\begin{cases} \dot{x} = & m\,a\,b_1\,y\,(1-x) - \gamma\,x \\ \\ \dot{y} = & b_2\,a\,(1-y)\,x - \mu\,y \end{cases} \tag{4.3}$$

With the preceding noations

$$\mathcal{F} = \begin{bmatrix} m\,a\,b_1\,y\,(1-x) \\ b_2\,a\,(1-y)\,x \end{bmatrix} \qquad \mathcal{V} = \begin{bmatrix} -\gamma\,x \\ -\mu\,y \end{bmatrix}$$

then

$$F = \begin{bmatrix} 0 & m\,a\,b_1 \\ b_2\,a & 0 \end{bmatrix} \qquad V = \begin{bmatrix} -\gamma\,x \\ -\mu\,y \end{bmatrix}$$

Therefore

$$F\,V^{-1} = \begin{bmatrix} 0 & \dfrac{m\,a\,b_1}{\mu} \\ \dfrac{b_2\,a}{\gamma} & 0 \end{bmatrix}$$

Then

$$\mathcal{R}_0^2 = \frac{m\, a^2\, b_1\, b_2}{\gamma\, \mu}$$

Chapter 5

Monotone systems in Epidemiology

5.1 Generalities

5.1.1 Introduction

We write this chapter on monotone system applied to epidemiology for many reasons. The first one is to have a self-contained lectures notes. The second reason is that this is rarely taught in the academic lectures, moreover there are no elementary references and results are scattered in the literature.
However it exists two excellent references at research level : [49, 44].

The following notes in this chapter are devoted to monotonous systems, a concept introduced by M. W. Hirsch in a series of founding articles [43, 45, 46, 47]. These systems appear quite often in biology.
We cannot speak of monotonous systems, without using the Perron-Frobenius theorem. The Perron-Frobenius theorem manifests a certain ubiquity in applied mathematics. We will see this further. In this chapter we will give the proofs of these fundamental theorems. The notion of $\mathcal{R}_0$ is based on this theorem.
A Metzler matrix is a matrix whose terms outside the diagonal are nonnegative. Economists (Arrow 1966) gave this name to this type of matrix because of their study by L. Metzler. These matrices have many applications in economics but also in all areas where we mode with compartmental systems. We can also speak of the ubiquity of Metzler's matrices. In fact there are two competing schools: Those who use the matrices of Metzler, still called quasi-positive matrices [84, 86, 92], and those using the opposite matrices, called $\mathcal{Z}$ -matrices [14, 94, 96]. The $\mathcal{Z}$-matrices contain the M-matrices for which there is an abundant literature.

The positive matrices have the characteristic of leaving invariant, as applications, the nonnegative orthant. In other words, A is positive if, and only if, $A\,\mathbb{R}^n_+ \subset \mathbb{R}^n_+$.
For discrete dynamic systems we will consider $x_{n+1} = A\,x_n$.
We must be careful not to confuse this notion of positive matrix with the notion of positive symmetric matrix. Here it will means that the entries of the matrix are nonnegative.
If we consider differential equations and if one looks for matrices such that the dynamic system $\dot{x} = A\,x$ leaves invariant the nonnegative orthant we obtain naturally Metzler matrices. There are a strong similarity between linear discrete systems and linear differential systems. We will exploit the dynamic properties of these linear differential systems to make the dictionary work in the other direction. In other words, A is nonnegative if, and only if, $A\,\mathbb{R}^n_+ \subset \ R^n_+$.

5.1.2 Generalities and Notations. Cones and Ordered relation

The ordered space $\mathbb{R}^n$

In a standard way if $x \in \mathbb{R}^n$ is a vector, we denote by x_i its i-th component.

Definition 5.1.1 *We define an order on $\mathbb{R}^n$ by $x \geq y$ if for any index i the inequality $x_i \geq y_i$ is satisfied*

It is easy to see that this relation is an order relation that makes $\mathbb{R}^n$ an ordered vector space. Note that this is partial order. In an ordered vector space of finite dimension, the nonnegative elements form a closed convex cone. This role is played, for the standard order, by the nonnegative orthant.

The following notation are now well recognized.
We write $\mathbb{R}^n_+$ the nonnegative orthant. We have the equivalence $x \geq y$ and $x - y \in R^n+$, especially

$$x \geq 0 \iff \text{for any index i we have} \quad x_i \geq 0$$

Notation $x > 0$ means $x \geq 0$ and $x \neq 0$
We will denote $x \gg 0$ if x is in the interior of $\mathbb{R}^n_+$, in other words

$$x \gg 0 \iff \text{for any index } \ i \quad x_i > 0$$

Remark 5.1.1 *These notations are now well admitted [14, 49, 86, 44], but it was not always the case. For example in some papers $x < y$ will means what we denote by $x \ll y$ [43, 85]. So when reading or citing the literature pay attention to the notations used*

Similarly, we extend these notations to matrices by assimilating the vector space of matrices $M(n,n),\mathbb{R})$ with $\mathbb{R}^{n^2}$. We write $A \geq B$ if for every couple of index (i,j) we have $a_{i,j} \geq b_{i,j}$ and we have the analog for $A > B$ and $A \gg B$.

For this order on $\mathbb{R}^n$ we define the closed interval

$$[[\mathbf{a},\mathbf{b}]] = \{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} \mid \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\} = [\mathbf{a_1},\mathbf{b_1}] \times \cdots \times [\mathbf{a_n},\mathbf{b_n}]$$

This notation must not to be confused with the notation for segment

$$[a,b] = \{ta + (1-t)b \ \mid \ 0 \leq t \leq 1\},$$

which is used in convexity.
In the same manner is defined the open interval

$$]]\mathbf{a},\mathbf{b}[[= \{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} \mid \mathbf{a} \ll \mathbf{x} \ll \mathbf{b}\} =]\mathbf{a_1},\mathbf{b_1}[\times \cdots \times]\mathbf{a_n},\mathbf{b_n}[$$

If E and F are subspaces of $\mathbb{R}^n$ we define

$$\mathbb{R}_+ E = \{\lambda\, x \ \mid \ \lambda \in \mathbb{R}_+ \quad x \in E\}$$

and

$$E + F = \{x + y \ \mid \ x \in E \quad y \in F\}$$

We will denote $\langle x \mid y \rangle$ the euclidean inner product of two vectors. If A is a matrix A^T will denote the transpose. If vectors of $\mathbb{R}^n$ are identified with column matrices $n \times 1$, then the inner product is expressed by $\langle x \mid y \rangle = x^T\, y$
We denote by e_i le i-th vector of the canonical basis of $\mathbb{R}^n$.

Cones and Order

The nonnegative orthant is $\mathbb{R}^n_+$ is a pointed convex cone (that means that 0 is in the cone).

Definition 5.1.2 (Cone)
A cone in a real vector space is a set K which satisfies

$$R^*_+\, K \subset K$$

In other words it is a set invariant by homotheties
A cone is said salient if it does not contain a pair of non zero opposite vectors, or equivalently does not contains a vectorial line.

$$K \bigcap -K = \{0\}$$

Definition 5.1.3 *A closed convex salient cone is a non empty convex set K, satisfying*

1. $K + K \subset K$
2. $\mathbb{R}_+ K \subset K$
3. $K \cap -K = \{0\}$

It is easy to show that such a set is convex.
In the remaining cone will means : closed convex salient cone

An order relation is said to be compatible with the structure of vector space if

$$\text{for any } z \quad x \leq y \Longrightarrow x + z \leq y + z \tag{5.1}$$

$$\text{for any } \lambda \geq 0 \quad x \leq 0 \Longrightarrow \lambda\, x \geq 0 \tag{5.2}$$

If moreover we have

$$x_n \leq y_n \Longrightarrow \lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n$$

we will say that the order is compatible with the topology. If these 3 properties are satisfied then we say that we have a ordered topological vector space.

Proposition 5.1.1
An order is compatible with the vector space if the set of elements ≥ 0 is a close convex salient cone. Reciprocally a cone define an order.
A topological vector space is ordered iff the set of elements ≥ 0 is a close convex salient cone

Orthant and faces

Definition 5.1.4
A subset F of the orthant is a face of the orthant if F is a cone and for any $x \in F$ the relation $0 \leq y \leq x$ implies $y \in F$

A face F is said proper if $F \neq \{0\}$ and $F \neq \mathbb{R}^n_+$.
It is easy to see that faces are

$$F_I = \{x \in \mathbb{R}^n_+ \mid x_i = 0 \text{ si } i \in I \subset [1, \cdots, n]\}$$

Face F_I is the convex cone generated by e_i for index in I.
The dimension of a face F is the dimension of the vector space generated by this face, in other words the vector space generated by the e_i for $i \in I$. As a face is a

convex cone, the generated vector space is also $F - F$. The faces $\{0\}$ and $\mathbb{R}^n_+$ are called the trivial faces. The faces of dimension 1 are the half-lines $\mathbb{R}_+ e_i$.
If x is an element of the orthant the set $\mathbb{R}_+[[0, x]]$ is the face generated by the element x.
We have two ways of defining a face. By a system of equations. This is what we choose here, with the notation F_I. If we denote $J = [1, n] \setminus I$, F_I is defined by the set of equations

$$x_j = 0 \text{ fort any } j \in J.$$

But if we consider F_I as the convex cone generated by some vectors of the canonical basis e_i for $i \in I$, this correspond to the parametric defintion :

$$x \in F_I \iff x = \sum_{i \in I} \lambda_i \, e_i \quad \lambda_i \geq 0$$

The positive orthant is an intersection of half-space hyperplanes. This is called a polyhedron. The faces are the faces of this polyhedron.

5.2 Monotone application and Monotone vector field

Definition 5.2.1
The application $f : R^n \longrightarrow R^n$ is said to monotone nondecreasing if for any pair (x, y) we have

$$x \leq y \Longrightarrow f(x) \leq f(y)$$

Application f is said strongly monotone if

$$x < y \Longrightarrow f(x) \ll f(y)$$

If F class $\mathcal{C}^1$ vector field (or an ODE) we can associate to the vector field the local flow $\phi_t(x_0)$, i.e., the value of the system

$$\begin{cases} \dot{x} & = F(x) \\ x(0) & = x_0 \end{cases}$$

at time t is defined en t.

Definition 5.2.2 (Monotone nondecreasing vector field)
A vector fiels is said monotone nondecreasing (monotone for short) if the associated flow $\phi_t(.)$is monotone nondecreasing i.e.,

$$x \leq y \Longrightarrow \phi_t(x) \leq \phi_t(y)$$

Definition 5.2.3 (Strongly Monotone nondecreasing vector field)
A vector fiels is said strongly monotone nondecreasing if the associated flow $\phi_t(.)$is strongly monotone nondecreasing i.e.,

$$x < y \Longrightarrow \phi_t(x) \ll \phi_t(y)$$

5.2.1 Monotone linear applications

Nonnegative matrices are characterized by an invariance property.

Proposition 5.2.1
Linear monotone applications correspond to nonnegative matrices.
A matrix is nonnegative, if and only if, it leaves invariant the nonnegative orthant $\mathbb{R}^n_+$.

Indeed, if $x \mapsto A\,x$ is monotone, then for all $x \leq y$ we have $A\,x \leq A\,y$. Equivalently, $y - x \geq 0$ results in $A\,(y - x) \geq 0$. A leaves invariant the nonnegative orthant. So for every vector of the canonical basis e_j, $A\,e_j \geq 0$ and consequently

$$A(e_i, e_j) = \langle A\,e_j \mid\ e_i \rangle \geq 0$$

Conversely if $A \geq 0$ then for all $x \geq 0$ we have $A\,x \geq 0$.

5.2.2 Metzler Matrices: Dynamical properties

We will show that the Metzler matrices leave "dynamically" invariant the nonnegative orthant and this characterizes them. More precisely we will study the linear systems $\dot{x} = Ax$ and search among these systems those which leave positively invariant the nonnegative orthant.

Theorem 5.2.1
The linear system $\dot{x} = Ax$ has the nonnegative orthant positively invariant iff A is Metzler

Proof
Sufficient : we prove that $e^{t\,A} \geq 0$, for any $t \geq 0$, if A is Metzler.
We have

$$e^{t\,A} = e^{-c\,t}\,e^{t(A+c\,I}$$

For c big enough $B = A + c\,I \geq 0$, then $e^{t(A+c\,I} \geq 0$. Indeed if $B \geq 0$ we have, from the result with nonnegative matrices :

$$\langle e^{t\,B}\,e_j \mid e_i\rangle = \sum_{k\geq 0} \frac{t^k}{k!}\,\langle B^k\,e_j \mid e_i\rangle \geq 0$$

Necessarily if $e^{t\,A} \geq 0$, for any $t \geq 0$, we will shiw that A is Metzler.
We have

$$A = \frac{d}{dt}(e^{t\,A})\Big|_{t=0} = \lim_{\substack{t\to 0\\ t>0}} \frac{e^{tA} - I}{t}$$

Therfeore for $i \neq j$

$$\langle A\,e_j \mid e_i\rangle = \lim_{\substack{t\to 0\\ t>0}} \frac{\langle e^{tA}\,e_j \mid e_i\rangle - \langle\, e_j \mid e_i\rangle}{t} = \lim_{\substack{t\to 0\\ t>0}} \frac{\langle e^{tA}\,e_j \mid e_i\rangle}{t} \geq 0$$

■

Proposition 5.2.2 *Linear system $\dot{x} = Ax + b$ leaves positively invariant the nonnegative orthant iff A is Metzler and $b \geq 0$*

Analogous proof as before.

This result is atrributed to ? Karlin by economy Nobel prize K. Arrow [9] and Bellman however never published by Karlin.

There is a discret analogue of this result. One wonders what are the matrices A, such that for the discrete system $x_{n+1} = A\,x_n$ leaves positively invariant the nonnegative orthant. It is clear that if we look for the matrices A which leave invariant the orthant $A\,\mathbb{R}^n + subset\mathbb{R}^n+$. It is immediate that these are the nonnegative matrices $A \geq 0$

5.2.3 Characterization of Hurwitz Metzler matrices

A book by Berman et Plemmons (1979) gives 50 equivalent condition for a Metzler matrix to be stable [14].
Actually this is a convenient abuse of language actually we means by stable a matrix such that $\dot{x} = A\,x$, asymptotically stable at the origin.
Here we will give some of the most important and prove them. In fact we will only need property 2, the others are classical and not expensive to demonstrate. The proofs given by [14] are generally lengthy algebraic proofs. Using the power of Lyapunov and LaSalle techniques gives easy, elegant and short proof.
We recall some definition for matrices

Definition 5.2.4
If we denote by spec(A) *the set of eigenvalues of A, i.e., the spectrum, we define the stability modulus, denoted by* $s(A)$ *the real number*

$$s(A) = \max_{\lambda \in \text{spec(A)}} Re(\lambda)$$

The spectral radius $\rho(A)$ *is*

$$\rho(A) = \max_{\lambda \in \text{spec(A)}} |\lambda|$$

Theorem 5.2.2
If A is Metzler, the the following assertions are equivalent

1. *Metzler matrix A is Hurwitz (asymptotically stable, i.e.,* $s(A) < 0$*)*
2. *Metzler matrix A is invertible and* $-A^{-1} \geq 0$
3. *If b is a vector such that* $b \gg 0$ *then there exists* $x \gg 0$ *such that* $Ax + b = 0$
4. *It exists* $c > 0$ *such that* $A\,c \ll 0$
5. *Il exists* $c \gg 0$ *such that* $A\,c \ll 0$

Proof

$(1 \Rightarrow 2)$:
Let any norm on $\mathbb{R}^n$. Since A is AS, we know that [48] ther exists a constant K such that for any t x_0 and for any $t \geq 0$ we have

$$\|e^{tA}\,x_0\| \leq K\,e^{s(A)\,t}\,x_0$$

This implies that the integral

$$\int_0^{+\infty} e^{tA}\,x_0\,dt$$

is normally convergent for any x_0.
We deduce the existence of $\int_0^{+\infty} e^{tA}\,dt$
Matrix being A Hurwitz we have $\lim_{t \to +\infty} e^{tA} = 0$.Since A is invertible

$$-A^{-1} = \int_0^{+\infty} e^{tA}\,dt = \left[A^{-1}e^{tA}\right]_0^{+\infty}$$

Using component (i,j) of $-A^{-1}$ given by $\langle -A^{-1}\, e_j \mid e_i \rangle$ we get

$$(-A^{-1})_{i,j} = \int_0^{+\infty} \langle e^{tA}\, e_j \mid e_i \rangle \, dt \geq 0$$

Indeed from 5.2.1 we have $e^{tA}\, e_j \geq 0$

$(2 \Rightarrow 3)$:
La solution de $Ax + b = 0$ est donn?e, si A est inversible, par $-A^{-1}b$. Comme $b \gg 0$ et $-A^{-1} \geq 0$ et qu'aucune ligne de $-A^{-1}$ ne peut ?tre identiquement nulle on en d?duit $x = -A^{-1}b \gg 0$

$(3 \Rightarrow 4)$:
WE choose $b \gg 0$ from 3, thers exists $c > 0$ such that $Ac + b = 0$ (we weaken the assertion) therfore $Ac = -b \ll 0$

$(4 \Rightarrow 5)$:
It is sufficient to perturb 4. Indeed let $\varepsilon > 0$ an $c_1 = c + \varepsilon \sum_{i=1}^n e_i \gg 0$.
Then $Ac_1 = Ac + \varepsilon \sum_{i=1}^n Ae_i$. By a continuity argument we can choose $\varepsilon > 0$ small enough such that $Ac_1 \ll 0$.

$(5 \Rightarrow 1)$:
We consider on the nonnegative orthant the ODE $\dot{x} = A^T x$. Choosing

$$V(x) = \langle c \mid x \rangle$$

Since $c \gg 0$ lfunction V is definite positive on $\mathbb{R}^n_+$.

$$\dot{V} = \langle c \mid Ax \rangle = \langle A^T c \mid x \rangle$$

This last quantity is zero iff $x = 0$. This prove the GAS of A^T on $\mathbb{R}^n_+$ by Lyapunov theorem. Since any initial condition x_0 can be written $x_0 = x_0^+ - x_0^-$ with x_0^+ and x_0^- in the nonnegative orthant we deduce that $e^{tA}x_0$ converges to the origin .This prove that A^T hence $A$is Hurwitz.

■

5.3 Perron-Frobenius Theorems

We will prove Perron's Theorem.
We state the weak form of Frobenius' s Theorem

Theorem 5.3.1
The spectral radius $\rho(A)$ of a nonnegative matrix A is an eigenvalue of A and there exists a corresponding nonnegative eigenvector.

In other words
If $A \geq 0$ $\exists v > 0$ such that $A\,v = \rho(A)\,v$
We get immediately the corollary

Theorem 5.3.2
The stability modulus $s(A)$ of a Metzler matrix A is an eigenvalue of A an there exists a corresponding nonnegative eigenvector.

In other words
If A is Metzler $\exists v > 0$ such that $A\,v = s(A)\,v$
To prove Perron, we need a fix point theorem, Brouwer's Theorem

Theorem 5.3.3 (Brouwer) *Any continuous application of a compact convex set $K \subset \mathbb{R}^n$ into K has a fix point.*

This theorem is still true when $\mathbb{R}^n$ is replaced by a Banach space: Schauder's Theorem

It has a generalization to locally convex topological vector spaces vectoriels topologiques localement convexes : Tychonoff-Kakutani 's Theorem

Perron-Frobenius Proof
Consider K

$$K = \{x \geq 0 \ \mid \ \|x\|_1 = 1 \ \text{ et } \rho(A)\,x \leq A\,x\}$$

Let v such that $A\,v = \lambda\,v$ with $|\lambda| = \rho(A)$. By dividing by the 1-norm, we can assume $\|v\|_1 = 1$.

$$\rho(A)\,|v| = |\rho(A)\,v| = |A\,v| \leq A\,|v|$$

$v \in K$, then K is non empty. Now it is easy to check that K is compact and convex.
If there exists $x \in K$ tel que $A\,x = 0$, we are finished (and $\rho(A) = 0$)
Otherwise we define a function on K

$$f(x) = \frac{1}{\|A\,x\|_1}\,A\,x$$

$$f(x) = \frac{1}{\|A\,x\|_1}\,A\,x$$

This function is continuous and

$$A\,f(x) = \frac{1}{\|A\,x\|_1}\,A\,A\,x \geq \frac{\rho(A)}{\|A\,x\|_1}\,A\,x = \rho(A)\,f(x)$$

In other words f send K in K, then f admit a fix point $y \in K$, by Brouwer 's Theorem

$$\frac{1}{\|A\,y\|_1}\,A\,y = y$$

This means that y is an eigenvector of A of eigenvalue $\|A\,y\|_1$.
Bur since $y \in K$, we have

$$\|A\,y\|_1\,y = A\,y \geq \rho(A)\,y$$

from $y > 0$, we deduce $\|A\,y\|_1 \geq \rho(A)$.
therefore $\rho(A) = \|A\,y\|_1$ ■

Perron's Theorem is for nonnegative matrices. Perron-Frobenius 's Theorem improve Perron's Theorem conclusion at the price of a supplementary hypothesis.To state Perron-Frobenius's Theorem we have to introduce another notion for matrices

5.4 Irreducible Matrices

This term was coined in 1912 by Frobenius.

Definition 5.4.1 *A square matrix $n \times n$, for $n \geq 2$, denoted $A = (a_{ij})$ is irreducible if for any proper subset of the set I of indices $\{1, \cdots, n\}$, there exists one index $j \in I$ and one index $i \notin I$ such that $a_{ij} \neq 0$. A matrix 1×1 is irreducible if it is nonzero.*

The geometrical interpretation is that A has no invariant subspace V_I of form $V_I = \{x \in \mathbb{R}^n \mid \text{ for any } j \in I \quad x_j = 0\}$. Indeed V_I is generated by the set linearly independent of vectors $\{e_j \mid j \in I\}$. This subspace will not be invariant if there exists a vector Ae_j for $j \in I$ which cannot be expressed as a linear combination of e_i for $i \in I$

This means there is no permutation P of coordinates such that in this new coordinates $P^{-1}AP$ is not in block form

$$P^{-1}AP = \begin{bmatrix} E & F \\ 0 & G \end{bmatrix}$$

where diagonal blocks are at least of dimension 1.
The irreducibility cannot be characterized by a beautiful graph theoretic interpretation

Definition 5.4.2 *A digraph (for directed graph) $G = (X, U)$ is a pair of n points $X = \{x_1, \cdots, x_n\}$ with a subset U of $X \times X$*
Elements of X are called nodes (also called vertices) of the graph. An element $(x, y) \in U$ is called an edge , x is the origin and y its end. It said that the edge leads x to y

A graph is a set of nodes with edges connecting some nodes. The edges are oriented

Definition 5.4.3 *A path is a sequence of edges $(u_1, \cdots, u_p)$ such that each u_i has for end the origin of u_{i+1}. We say that the origin of u_1 is connected to the end of u_p*
A graph is strongly connected if any pair (x, y) of vertices there is path which leads x to y

Associating a graph to a matrix

Definition 5.4.4 *If $A = (a_{ij})$ is a $n \times n$ matrix, we consider the graph with n vertices $X = \{1, \cdots, n\}$. An edge leads vertice i to vertice j if $a_{ji} \neq 0$. We say that a_{ji} is the weight of the edge (i, j) .*
Conversely to any n graph we can associate a matrix $n \times n$, where $a_{ij} = 1$ if there is an edge from i to j and $a_{ij} = 0$ otherwise.

You will note the inversion of indices, this is to stick to the representation of compartmental model. In our definition a_{ji} is the flow from i to j. It is clear that the irreducibility of A is equivalent to the irreducibility of A^T.

Remark 5.4.1
The diagonal terms has no role in irreducibility

Theorem 5.4.1 *:*
A irreducible iff its associated graph $G(A)$ is strongly connected

Proof
Condition is necessary

Assume A irreducible. Let i a vertice. We define by I the set of vertices, different of i, accessible by a path from i, i.e. the set of vertices $j \neq i$ such that there is path which leads i to j.

The set I is non empty. Indeed let J the complement set of the singleton $\{i\}$. Since A is irreducible there exists $k \notin J$ and $j \in J$ such that $a_{jk} \neq 0$. But with the definition of J this means that there exists $j \neq i$ such that $a_{ji} \neq 0$. There exists a path from i to j.

We assume, argument by absurdity, that $I \neq \{1, \cdots, n\}$. By irreducibility of A there exists $j \in I$ and t $k \notin I$ such that $a_{kj} \neq 0$. Hence we have a path leading from j to k. Since $j \in I$, j is accessible from i, hence ki is accessible from i. A contradiction.

Sufficient condition. Assume again by an argument of absurdity that the associated graph is strongly connected and A reducible. Then there exists a proper subset of indices I, such that if J is its complement, we have $a_{ji} = 0$ for any $i \in I$ and any $j \in J$. We have simply taken the negation of the property of irreducibility. We choose and index $i \in I$ and an index $j \in J$. This is possible since I is proper. Now we know that there is path from i to j. Then there exists indices $\{k_1, \cdots, k_p\}$ such that the following entries are non zero

$$a_{j,k_1}, a_{k_1,k_2}, \cdots, a_{k_p,i}$$

With hypothesis on I and J we deduce that since $a_{k_p,i} \neq 0$, $k_p \notin J$, soit $k_p \in I$. But if $k_p \in I$ the same argument applied to a_{k_{p-1},k_p} proves that $k_{p-1} \in I$. A finite induction argument proves $j \in I$, a contradiction.

■

Corollary 5.4.1
If a nonnegative matrix A is irreducible, then for any pair of indices (i, j), if $i \neq j$ there exists $k \in \mathbb{N}$ such that

$$A^k(j, i) \neq 0$$

Conversely if $A^k(i, j) \neq 0$ there exists a path of length k leading from i to j.

Proof
Consider $A^2(i, j)$

$$A^2(j, i) = \sum_{k=1}^{n} a_{j,k}\, a_{k,i}$$

a_{jk} is non zero if there exists an edge from k to j, entry $a_{k,i}$ is non zero if there is an edge from i to k. Since the sum is a sum of nonnegative terms, $A^2(j,i)$ will be non zero if there exists a path of length 2 from i to k.
By $A^k(j,i)$ is non zero if there exists a path of length k leading from i to j. Irreducibility implying strong connectedness, the proof is finished. ■

An reducible matrix A can be transformed in block form

$$P^T A P = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ 0 & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{pp} \end{bmatrix}$$

Where diagonal A_{ii} are irreducible matrices and P a permutation matrix.
Apply reduction process by induction

Proposition 5.4.1
a nonnegative matrix A is irreducible iff A leaves invariant no nontrivial face of the nonnegative orthant.

left as an exercise.

5.4.1 Irreducible Metzler Matrices

We will characterize by a dynamic property the irreducibility of Metzler matrices. An linear ODE with Metzler matrix let positively invariant the nonnegative orthant. What is happening at the border, in other words on the faces of the orthant?

Proposition 5.4.2 *:*
If A is an irreducible Metzler matrix, no trajectory can remain in a face of the nonnegative orthant.

More precisely
A is an irreducible Metzler matrix then for any $t > 0$ we have $e^{tA} \gg 0$.
Particularly if $x_0 > 0$ then $e^{tA} x_0 \gg 0$.

Proof
We will begin to bring back to a nonnegative matrix. For $c \geq 0$ large enough matrix $A + c\,I$ is nonnegative. Since

$$e^{t\,A} = e^{-c\,t}\, e^{t(cI+A)}$$

and $e^{-c\,t} > 0$, it is sufficient to prove the proposition for $cI + A \geq 0$. We already know that diagonal terms does not count in irreducibility. Then $A + c\,I$ is irreducible if A is .
We can assume that $A \geq 0$. Then we have, by analyticity

$$\langle e^{t\,A}\,e_j|\,e_i\rangle = \sum_{k\geq 0} \frac{t^k}{k!}\,\langle A^k\,e_j|\,e_i\rangle$$

This series is a sum of nonnegative terms. The sum will be positive if one term is positive. Since the matrix is irreducible nonnegative there exists from corollary (5.4.1) a natural number k such that

$$A^k(j,i) = \langle A^k\,e_j|\,e_i\rangle \neq 0$$

This ends the proof of $e^{t\,A} \gg 0$.
Therefore if x is a vector $x > 0$ then $e^{t\,A}\,x \gg 0$. Any positive trajectory starting from the orthant is immediately in the interior. This ends the proof.

■

This shows that e^{tA}, corresponding to irreducible Metzler matrices , is a strongly monotone application.

Proposition 5.4.3
Linear system $\dot{x} = A\,x$, is strongly monotone, i.e., the linear application e^{tA} is strongly monotone iff A is an irreducible Metzler matrix.

Proof
Condition is sufficient : this is the preceding proposition.
Condition is necessary. We know that A is necessarily Metzler since $e^{tA} \gg 0 \geq 0$. This is 5.2.1. We will give a contrapositive proof. Again writing

$$e^{tA} = e^{-st}\,e^{((sI+A)}$$

we can assume $e^{sI+A} \gg 0$. Then we assume that $A \geq 0$. Assuming A reducible, it exists a proper face F, invariant by A, i.e.,

$$A\,F \subset F$$

By induction we observe that A^n leave F invariant. Since F is a cone, we have for any $t \geq 0$

$$F + t\,A\,F + \frac{t^2}{2!}\,A^2\,F + \cdots + \frac{t^n}{n!}\,A^2\,F \subset F$$

Face F is a closed set, a we just prove that $e^{tA}\,F \subset F$. A contradiction with $e^{tA} \gg 0$.

■

5.4.2 Perron-Frobenius

We will give another characterization of irreducility for Metzler matrices

Proposition 5.4.4 (Irreducible Metzler Matrices)
The Metzler matrix A is irreducible iff for any vector $x > 0$ in a face F of $\mathbb{R}^n_+$, where F is defined by

$$F = \{x \geq 0 \mid i \in I \;\; \langle e_i \mid x\rangle = 0\},$$

there exists an index $i \in I$ such that $\langle e_i \mid A\,x\rangle > 0$.

This means, geometrically, that for any face of the nonnegative orthant, the vector field associated to A , for any point of F is never tangent to F.

Proof

The condition is necessary

Again we can replace A by $A + \lambda\, I_n$, wher I_n is the identity matrix, for λ large enough. Indeed if there exists i such that $\langle e_i \mid x\rangle = 0$ and $\langle e_i \mid A\,x\rangle > 0$ this is equivalent to $\langle e_i \mid x\rangle = 0$ and $\langle e_i \mid (A + \lambda I_n\,)x\rangle > 0$. Then we will assume $A \geq 0$. We will give a contrapositive proof.

Assume i that for any $i \in I$ $\langle e_i \mid x\rangle = 0$ we have $\langle e_i \mid A\,x\rangle = 0$.
Let $F_x = \mathbb{R}^+\,[[0, x]]$ the face generated by par x. (Exercise show that this is the smallest face containing x)
Since $A \geq 0$ we have $A\,F_x = \mathbb{R}^+\,[0, Ax]$. F_x is characterized by a set if indices I.
We have $F_x = \{x \geq 0 \mid \langle e_i \mid x\rangle = 0\}$. For these indices we have $\langle e_i \mid A\,x\rangle = 0$.
Therefore $A\,F_x \subset F_x$. A face is invariant by A, then the matrix is not irreducuble.
Necessary condition.

Again by a contrapositive argument.

If A is reducible, there exists a face, which can be written F_x, such that $A\,F_x \subset F_x$. For any index such that $\langle e_i \mid x\rangle = 0$ we then have $\langle e_i \mid A\,x\rangle = 0$.

■

Theorem 5.4.2 (Characterization)
A Metzler matrix is irreducible iff one of the following assertions is satisfied

1. *An eigenvector of A cannot belong to a face of the nonnegative orthant.*
2. *Matrix A has exactly one eigenvector $v \gg 0$ (up to a positive multiplicative factor). It is associated to the stability modulus*
3. *Condition $x > 0$ with $A\,x \leq \alpha\,x$ implies $x \gg 0$.*
4. *$(I + A)^{n-1} \gg 0$.*

Proof of condition (1) and (3)

We prove condition **(1)**. Condition **(1)** is sufficient by a contrapositive argument. Assume A leaves invariant a subspace V_I generated by a face. We consider the restriction $A|V$ of application A to V. This is again a Metzler Matrix.Therefore, Perron's Theorem can be applied to ? $A|V$: there exists an eigenvalue $v > 0$ of $A|V$, therefore of A, in V_I, hence $v \in \partial R^n_+$.

Condition **(1)** is necessary. Let $v \in \partial R^n_+$ an eigenvector of A and a canonical basis vector e_i such that $\langle v|e_i\rangle = 0$. We j-have $e^{t\,A}\,v = e^{t\lambda}\,v$. Then $\varphi(t) = \langle e^{t\,A}\,v|e_i\rangle = 0$. Function φ is differentiable and we have

$$\varphi'(t) = \langle A\,e^{t\,A}\,v|e_i\rangle = \langle \lambda\,e^{t\,\lambda}\,v|e_i\rangle = 0$$

Therefore $\varphi'(0) = \langle A\ v|\ e_i\rangle = 0$, this proves that A is reducible from proposition (5.4.4).
we will now prove condition **(3)**. If A is reducible there is an invariant face and an eigenvector v in this face : this is obtained by applying Frobenius Theorem to the restriction of A to F. Then there exists $v > 0$ with $A\,v = \lambda\,v \leq s(A)\,v$ and however v does not satisfies $v \gg 0$.
Conversely for $A + \tau\,I \leq 0$, for any x satisfying the inequality , the face $F_x = \mathbb{R}_+\,[[0, x]]$ generated by x satisfies $A\,F_x \subset F_x$. Since no face can be positively invariant by A, we deduce $x \gg 0$.
end of proof condition (1) and (3)
To prove condition (**2**) we need some technical results

Lemma 5.4.1
If the nonnegative matrix A, has two eigenvectors in the interior of the nonnegative orthant, the associated eigenvalues are equal and there exists an eigenvector on the boundary of the orthant.

Proof

We need some technical results on a generalization of the infinity norm $\|\ \|_\infty$.

Proposition 5.4.5 (Weighted infinity norm) *:*
To any vector $v \gg 0$, we associate the norm defined by

$$\|x\|_v = \max_i \frac{|x_i|}{v_i}$$

This is clearly a norm on $\mathbb{R}^n$. If we denote by $|x|$ the vector whose components are $|x_i|$, we have

$$\|x\|_v = \inf_{v - \frac{|x|}{\lambda} \geq 0} \lambda = \inf_{t \geq 0 \; |x| \leq t\, u} t$$

The unit ball for $\| \; \|_v$ is $B_v = [[-v, v]]$. The unit sphere is a polyhedron., with faces parallel to the faces of the nonnegative orthant.
The vector $\frac{x}{\|x\|_v}$ belongs to the unit sphere as is v. This implies that for any $x > 0$ we have
Important remark

$$v - \frac{x}{\|x\|_v} \in \partial \mathbb{R}^n_+$$

It also can be checked by considering the components.
We can now prove the lemma

Let $A\, v_1 = \lambda_1\, v_1$ and $A\, v_2 = \lambda_2\, v_2$ with $v_1 \gg 0$ and $v_2 \gg 0$.
Since $A \geq 0$ eigenvalues λ_1 and λ_2 are nonnegative.
Assume $\lambda_2 \geq \lambda_1 \geq 0$. Let

$$v_3 = v_1 - \frac{1}{\|v_2\|_{v_1}}\, v_2$$

We have seen in the properties of the weighted norms that $v_3 \in \partial \mathbb{R}^n_+$. Therefore $A\, v_3 \geq 0$.
$A\, v_3 = \lambda_1\, v_1 - \lambda_2\, \frac{1}{\|v_2\|_{v_1}}\, v_2 \geq 0.$

If $\lambda_1 = 0$, then $\lambda_2 = 0$, we are finished
Otherwise

$$A\, v_3 = \lambda_1 \left[v_1 - \frac{\lambda_2}{\lambda_1}\, \frac{1}{\|v_2\|_{v_1}}\, v_2 \right] \geq 0$$

With the definition of $\|v_2\|_{v_1}$, necessarily we have $\lambda_1 \geq \lambda_2$. (Compute the coordinates of$A\, v_3$, look for the index where the maximum is reached, and use $A\, v_3 \geq 0$
This prove $\lambda_1 = \lambda_2$ with $v_3 \in \partial \mathbb{R}^n_+$ eigenvalue of A.

■

Back to Proof of Theorem (5.4.2)

Condition **(2)** of Theorem (5.4.2) is sufficient : assume A is nonnegative irreducible, it admits an eigenvector in the orthant from Perron's Theorem and with condition (**1**), already proved, of the Theorem, it cannot be in the boundary of the orthant. It is necessarily unique from the preceding lemma.
Conversely assume A has a exactly one eigenvector $v \gg 0$, $A\,v = \lambda\,v$. Since $A \geq 0$ we have $\lambda \geq 0$. By Perron's Theorem we have a nonnegative eigenvector x such that $A\,x = \rho(A)\,x$. If we consider $v - \dfrac{x}{\|x\|_v} \in \partial\mathbb{R}^n_+$, then

$$A\left(v - \frac{x}{\|x\|_v}\right) = \lambda\,v - \rho(A)\,\frac{x}{\|x\|_v} \geq 0.$$

Let i_0 such that $\dfrac{x_{i_0}}{v_{i_0}} = \max_i \dfrac{x_i}{v_i} = \|x\|_v$, then

$$\lambda\,v_{i_0} - \rho(A)\,\frac{x_{i_0}}{\left(\dfrac{x_{i_0}}{v_{i_0}}\right)} = (\lambda - \rho(A))\,v_{i_0},$$

which implies $\lambda \geq \rho(A)$ hence $\lambda = \rho(A)$. This proves that x and v are eigenvectors for the same eigenvalue $\rho(A)$. But $t\,x + (1-t)\,v$ are also eigenvectors for $\rho(A)$ and for t small enough be in the interior of the nonnegative orthant.
We have two eigenvectors $\gg 0$. a contradiction. Necessarily $x = v$

Condition **(4)** of Theorem (5.4.2). If A is reducible it exists $v > 0$, $v \in \partial\mathbb{R}^n_+$ such that $A\,v = \lambda\,v$. We have $(I + A)^{n-1} = (1 + \lambda)^{n-1}\,v$ in the boundary of the nonnegative orthant. Conversely if A is irreducible, let $v > 0$ in the boundary of the orthant. Then $(I + A)\,v \notin F_v$. The set $(I + A)\,F_v$ is necessarily in a face of dimension strictly greater than the dimension of F_v. By a finite induction $(I + A)^{n-1}\,v \gg 0$.
Therefore for any vector of the canonical basis e_i we have $(I + A^{n-1})\,e_i \gg 0$, which proves

$$(I + A^{n-1}) \gg 0.$$

■

Theorem 5.4.2 contain the following theorem

Theorem 5.4.3 (Perron-Frobenius)
Let A be an irreducible nonnegative matrix, then the spectral radius is a simple eigenvalue of A and its associated eigenvector v is positive. This vector is unique (up to a positive multiplicative factor)

In other words
If $A \geq 0$ irreducible then $\exists\, v \gg 0$ such that $A\,v = \rho(A)\,v$
We get immediately

Theorem 5.4.4 (Perron-Frobenius)
Let A an irreducible Metzler matrix, then the stability modulus is a simple eigenvalue of A and its associated eigenvector v is positive. Conversely A has exactly one eigenvector $v \gg 0$ (up to a positive multiplicative factor). It is associated to the stability modulus

If A is an irreducible Metzler matrix then $\exists\, v \gg 0$ such that $A\,v = s(A)\,v$

5.4.3 Stability modulus and order

The stability modulus is an increasing function on the set of Metzler matrices.

Theorem 5.4.5

1. *If theres exists $v \gg 0$ such that $A\,v \leq \beta\,v$ then $s(A) \leq \beta$.*
2. *If moreover A is irreducible then if $v > 0$ and $A\,v < \beta\,v$ imply $s(A) < \beta$. Actually we have necessarily $v \gg 0$.*
3. *If it exists $v > 0$ such that $\alpha\,v \leq A\,v$ then $\alpha < s(A)$.*

This theorem, for its nonnegative version and spectral radius is proved in [14] in a algebraic way (th?or?mes 2.1.11, 1.3.34 et 1.3.35). We will give a short proof using Lyapaunov and LaSalle.

Proof
Let $v \gg 0$ such that $A\,v \leq \beta\,v$. This is equivalent to $(A - \beta\,I)\,v \leq 0$.
Considering on the nonnegative orthant the definite positive function $V(x) = \langle c|x\rangle$ it appears that system $\dot{x} = (A - \beta\,I_n)^T\,x$ is stable. The trajectories starting from a nonnegative point are bounded. taking the proof argument of 5.2.2, i.e., the nonnegative orthant generate $\mathbb{R}^n$, it is clear that all trajectories are bounded. This linear system is not unstable hence its stability modulus in non positive, in other words

$$s(A - \beta\,I_n) = s(A) - \beta \leq 0.$$

If moreover A is irreducible, and $v > 0$ we consider the same function V. From 5.4.2 property 3, A irreducible and $A\,v < \beta\,v$ implies $v \gg 0$. V is definite positive. We look for the greatest invariant set E in $\mathcal{L}$ defined by par $\dot{V} = 0$, i.e.,

$$\mathcal{L} = \{x \geq 0 \;\; |\langle (A - \beta\,I_n)\,v|x\rangle = 0\}$$

Let $x \in E$ then any trajectory from x denote $\phi_t(x)$ stays in $\mathcal{L}$. Therefore for any $t > 0$

$$\langle (A - \beta\, I_n)\, v | \phi_t(x) \rangle = 0$$

then

$$\frac{d}{dt} \langle (A - \beta\, I_n)\, v | x \rangle = \langle (A - \beta\, I_n)\, v | \dot{x} \rangle = \langle (A - \beta\, I_n)^2\, v |\ x \rangle = 0$$

By induction we prove that for any k, if $x \in E$, we have $\langle (A - \beta\, I_n)^k\, v |\ x \rangle = 0$. This implies by lemma (5.4.2) that

$$\langle \exp(t\, (A - \beta\, I_n)\, v |\ x \rangle = 0$$

Since $(A - \beta\, I_n)$ is irreducible $\exp(t\, (A - \beta\, I_n) \gg 0$, hence $\exp(t\, (A - \beta\, I_n)\, v \gg 0$, which in turn implies that $x = 0$. Then the greatest invariant set contained in $\mathcal{L}$ is $\{0\}$. By LaSalle 's invvaraint principle $(A - \beta, I_n$ is Hurwitz. Which proves $s(A - \beta\, I_n) < 0$
E is simply Since $A - \beta\, I_n$ is irreducible, no face can be invariant, since $v \gg 0$ we have $(A - \beta\, I_n)\, v \ll 0$. This means that $\dot{V}$ is negative definite. By Lyapunov, the origin is GAS in the nonnegative orthant.
This proves that the origin is GAS, hence

$$s(A - \beta\, I) = s(A) - \beta < 0.$$

For the other inequalities if $v \gg 0$ arguments are identical. We simply consider the system $\dot{x} = (\alpha\, I_n - A)\, x$. We prove that this system is stable by an analogous argument, which proves $\alpha \leq s(A)$ Hurwitz. When $v > 0$ There is, here, a difference since we cannot ascertain that $v \gg 0$. We must use LaSalle's principle. Let consider the greatest invariant set E contained in $\mathcal{L}$ such that

$$\mathcal{L} = \{x \in \mathbb{R}^n_+ \mid \langle (\alpha\, I_n - A))\, v | x \rangle = 0$$

By a similar argument as before, we obtain if $x \in E$, for any $t \geq 0$

$$\langle \exp t\, (\alpha\, I_n - A))\, v | x \rangle = 0$$

Since $(-\alpha\, I_n + A)$ is Metzler a matrix (note the sign change), $\exp t\, (-\alpha\, I_n + A)) \geq 0$, this matrix is invertible, with $v > 0$ it implies that $\exp t\, (-\alpha\, I_n + A))\, v \gg 0$, hence

$$\langle \exp t\, (\alpha\, I_n - A))\, v | x \rangle = 0$$

iff $x = 0$. Then the origin is attractive and because we are with linear systems, the matrix $(\alpha\, I_n - A)$ is Hurwitz, implying $\alpha \leq s(A)$

■

Theorem 5.4.6
Let A and B two Metzler matrices

1. *If $A \leq B$ then $s(A) \leq s(B)$*
2. *If A is irreducible $A < B$ implies $s(A) < s(B)$.*

Proof
Let $v > 0$ such that $A\,v = s(A)\,v$. Then we have $A\,v = s(A)\,v \leq B\,v$. From Theorem (5.4.5 3), we deduce $s(A) \leq s(B)$.

If A is irreducible, there exist $v \gg 0$ such that $A\,v = s(A)\,v$. We have $A < B$, which implies $A\,v < B\,v$, let $s(A)\,v < B\,v$. If A is irreducible $A \leq B$, B is irreducible. We conclude with (5.4.5 4).

■

Corollary 5.4.2
If B is a principal matrix of A then $s(B) \leq s(A)$
If a Metzler matrix is Hurwitz its diagonal is negative.

Proof as an exercise.

5.5 Characterization of Monotone Dynamical Systems

We will see that actually the Metzler matrices are the infinitesimal version of the monotone vector fields

Definition 5.5.1 (Kamke-Muller condition)
The vector field F is said of type K if for any i and for any pair (x, y) such that $x \leq y$ and $x_i = y_i$ we have $F_i(x) \leq F_i(y)$.

This definition extends to non-autonomous vector fields $F(t, x)$. We consider the associated flow $\phi_{t,t_0}(x_0)$ pour $t \geq t_0$, associated to non-autonomous ODE.

$$\begin{cases} \dot{x} & = F(t, x) \\ x(t_0) & = x_0 \end{cases}$$

Properties considered in the assertions must be true for any t. We have the following theorem

Theorem 5.5.1
Let F be a class $\mathcal{C}^1$ vector field on a convex open set Ω. We use the order from the nonnegative orthant. Then the following are equivalent

1. *F is monotone*
2. *F is of type K*
3. *F is such that, for any $x \in \Omega$, the Jacobian $DF(x)$, computed at x, is a Metzler matrix.*

A vector field satisfying condition (3) is said to be cooperative. We assume Ω convex, Actually it is sufficient to have a weaker condition. We say that Ω is order convex : if for any pair x and y in Ω satisfying $x \leq y$ then $t\,x + (1-t)\,y \in \Omega$ for any $t \in [0,1]$.

Proof

$\boxed{1 \Rightarrow 2}$

Let $x \leq y$ et $x_i = y_i$, by hypothesis, for any $t \geq 0$: $\phi_t(x) \leq \phi_t(y)$. Let e_i the i-th vector of the canonical basis. We have

$$\langle e_i \mid \phi_t(y)\rangle - \langle e_i \mid \phi_t(y)\rangle \geq 0.$$

Since $\langle e_i \mid \phi_0(y)\rangle - \langle e_i \mid \phi_0(y)\rangle = y_i - x_i = 0$, we deduce

$$\frac{d}{dt}\left[\langle e_i \mid \phi_t(y)\rangle - \langle e_i \mid \phi_t(y)\rangle\right]_{|t=0} = F_i(y) - F_i(x) \geq 0$$

This proves $1 \Rightarrow 2$.

$\boxed{2 \Rightarrow 3}$

We have $x + t\,e_j \geq x$ for any $t > 0$ and the i-components of $x + t\,e_j$ and x for $i \neq j$ are equal. If F is of type K, then if $i \neq j$, we have $F_i(x + t\,e_j) - F_i(x) \geq 0$ therefore $i \neq j$

$$\lim_{\substack{t \to \infty \\ t>0}} \frac{F_i(x + t\,e_j) - F_i(x)}{t} = \frac{\partial F_i}{\partial x_j}(x) \geq 0$$

Matrix $DF(x)$ is Metzler

$\boxed{3 \Rightarrow 2}$

Let $x \leq y$, such that $x_i = y_i$. Since Ω is order-convex for $s \in [0,1]$ we have $(1-s)\,x + s\,y \in \Omega$. Therefore

$$F_i(y) - F_i(x) = \left[\left(\int_0^1 DF((1-s)\,x + s\,y)\,ds \right) (y-x) \right]_i$$
$$= \sum_{j \neq i} \int_0^1 \frac{\partial F_i}{\partial x_j}((1-s)\,x + s\,y)\,ds\,(y_j - x_j) \leq 0$$

$\boxed{2 \Rightarrow 1}$

Let $x \leq y$ and assume $\phi_t(x)$ and $\phi_t(y)$ are defined for $t > 0$. We wont to get $\phi_t(x) \leq \phi_t(y)$.
Let $v \gg 0$. For example $v = e_1 + \cdots e_n$ and consider the ODE

$$\dot{x} = F(x) + \varepsilon\, v$$

We denote by $\phi_t^\varepsilon(\ )$ the associated flow. We know that for $\varepsilon > 0$ small enough, the flow $\phi_s^\varepsilon(y + \varepsilon\, v)$ will be defined on $[0, t]$. Furthermore $\phi_s^\varepsilon(y + \varepsilon\, v)$ uniformly converges on $[0, t]$. See lemma 3.1, chap 1 [37].

We will show that for ε small enough, for any $s \in [0, t]$ we have $\phi_s(x) \ll \phi_s^\varepsilon(y + \varepsilon\, v)$

We have $\phi_0(x) \ll \phi_0^\varepsilon(y + \varepsilon\, v)$, therefore this inequality is still satisfied, by a continuity argument, for s small enough. We will use an absurdity argument. If this were not true, it exists $s_0 > 0$ and an index i such that

$$\langle e_i \mid \phi_{s_0}(x) \rangle = \langle e_i \mid \phi_{s_0}^\varepsilon(y + \varepsilon\, v) \rangle,$$

with $\phi_{s_0}(x) \leq \phi_{s_0}^\varepsilon(y + \varepsilon\, v)$ and for any $s < s_0$

$$\phi_s(x) \leq \phi_s^\varepsilon(y + \varepsilon\, v)$$

Which implies for $0 \leq s < s_0$

$$\langle e_i \mid \phi_s(x) \rangle \leq \langle e_i \mid \phi_s^\varepsilon(y + \varepsilon\, v) \rangle$$

Consequently

$$\frac{d}{ds} \langle e_i \mid \phi_s^\varepsilon(y + \varepsilon\, v) - \phi_s(x) \rangle|_{s=s_0} = \lim_{\substack{s \to s_0 \\ s < s_0}} \frac{1}{s - s_0} \langle e_i | \, (\phi_s^\varepsilon(y + \varepsilon\, v) - \phi_s(x)) \rangle$$
$$= F_i(\phi_{s_0}^\varepsilon(y + \varepsilon\, v)) + \varepsilon\, v_i - F_i(\phi_{s_0}(x))$$
$$\leq 0$$

We deduce

$$F_i(\phi^{\varepsilon}_{s_0}(y+\varepsilon\, v)) < F_i(\phi^{\varepsilon}_{s_0}(y+\varepsilon\, v)) + \varepsilon\, v_i \leq F_i(\phi_{s_0}(x))$$

But in the other hand since F is of type K, with the hypotheses in s_0, we have

$$F_i(\phi_{s_0}(x)) \leq F_i(\phi^{\varepsilon}_{s_0}(y+\varepsilon\, v))$$

A contradiction which proves our claim, namely

$$\phi_s(x) \ll \phi^{\varepsilon}_s(y+\varepsilon\, v),$$

for any $s \in [0,t]$. Since $\phi^{\varepsilon}_t(y+\varepsilon\, v) \to \phi_t(y)$ when $\varepsilon \to 0$, by going to the limit we get

$$\phi_t(x) \leq \phi_t(y)$$

This ends the proof

■

Remark 5.5.1 *Theorem is still true for non-autonomous systems.*

Remark 5.5.2 *Let F a class $\mathcal{C}^1$, monotone vector fiels. We denote by $\prec$ any of the relations $\leq$, $<$ and $\ll$. If i $x \prec y$ then $\phi_t(x) \prec \phi_t(y)$.*

Proof
If $x \leq y$ we know that $\phi_t(x) \leq \phi_t(y)$ by definition. Since $\phi_t()$ is a diffeomorphism, it is bijection therefore $x < y$ implies $\phi_t(x) \neq \phi_t(y$ and consequently $\phi_t(x) < \phi_t(y$. If $x \ll y$ since we have $\phi_t()$ monotone,the interval

$$[x,y] = \{z \mid x \leq z \leq y\}$$

is sent in $[\phi_t(x), \phi_t(y)]$. Since $x \ll y$ the set $[x,y]$ has is interior empty and therefore t $[\phi_t(x), \phi_t(y)]$ also, since $\phi_t()$is a diffeomorphism. we deduce $\phi_t(x) \ll \phi_t(y)$.

■

Corollary 5.5.1
A monotone vector field, with the origin as a fix point leaves the nonnegative orthant and its interior positively invariant

Proof
It is sufficient to remark that if $x \geq 0$ then $\phi_t(x) \geq \phi_t(0) = 0$.
Int he same manner $x \gg 0$ implies $\phi_t(x) \gg \phi_t(0) = 0$.

■

The following proposition will extend to monotone linear non-autonomous vector fields.

Proposition 5.5.1 *We consider a non-autonomous linear equation, where for any $t \geq t_0$ matrix $A(t)$ is Metzler.*

$$\begin{cases} \dot{x} & = A(t)\, x \\ x(t_0) & = x_0 \end{cases}$$

Then the nonnegative orthant and its interior are positively invariant by the flow associated to the ODE.

Proof
The matrix $A(t)$ is Metzler, we claim that $A(t)\, x$ is of type K.
Indeed if $x_i = y_i$ et $x \leq y$ then

$$\langle e_i | A(t)\, (y - x) \rangle \geq 0$$

therefore

$$\langle e_i | A(t)\, y \rangle \geq \langle e_i | A(t)\, x \rangle$$

The vector field (non-autonomous) is monotone: from the preceding remark if $x > 0$ we have a$\phi_{t,t_0}(x) > \phi_{t,t_0}(0) = 0$. This proves the positive invariance of the nonnegative orthant. Moreover if $x \gg 0$ then $\phi_{t,t_0}(x) \gg 0$, this proves the positive invariance of the interior of nonnegative orthant.

■

Remark 5.5.3 *We can now prove directly* $3 \Rightarrow 1$

5.6 Strongly monotone vector fields

Monotone Linear vector field are associated to Metzler matrices. We will see that Strongly Monotone Linear vector field are associated to irreducible Metzler matrices Indeed we see A irreducible Metzleri, iff $e^{t\,A} \gg 0$. If $x < y$, or equivalently $y - x > 0$ with $e^{t\,A} \gg 0$ implies $e^{t\,A}\, (y - x) \gg 0$.

5.6.1 Linear vector fields strongly monotone

We prove the result for the linear non-autonomous vector fields.

Theorem 5.6.1
Let $R(t, t_0)$ the fundamental solution of the following ODE

$$\left\{\begin{array}{ll} \dot{X} & = A(t)\,X \\ X(t_0) & = I \end{array}\right.$$

We assume that $A(t)$ is Metzler for any $t \geq t_0$ that it exists a tile $s \in [t_0, t_1]$ such that $A(s)$ is irreducible.
Then $R(t_1, t_0) \gg 0$ For any $t_1 > t_0$.

Proof
By an absurdity argument. Assume it is false for a time $t_1 \geq t_0$. Then it exists $x > 0$ such that $R(t_1, t_0)\,x \in \partial\mathbb{R}^n_+$. Since $R(t_1, s)$ is invertible and nonnegative for any $s \leq t_1$ and since we have the one parameter group relation we have $R(t_1, s)\,R(s, t_0) = R(t_1, t_0)$. We see that $R(s, t_0)$ cannot be strongly positive and $R(s, t_0)\,x$ is in the boundary of the nonnegative orthant : The interior of the orthant is invariant. Since $A(s)$ is irreducible, there exists e_i such that $\langle e_i | R(s, t_0)\,x\rangle = 0$ and such that $\langle e_i | A(s)\,R(s, t_0)\,x\rangle > 0$. But since $R(t, t_0) \geq 0$, the function $\varphi(t) = \langle e_i | R(t, t_0)\,x\rangle$ is nonnegative for any $t \in [t_0, t_1]$. By hypothesis $\varphi(s) = 0$. It is a minimum. Therefore $\varphi'(s) = 0$, but in the other hand, also by hypothesis $\varphi'(s) = \langle e_i | A(s)\,R(s, t_0)\,x\rangle > 0$. A contradiction.

■

Corollary 5.6.1
Let the linear equation $\dot{x} = A(t)\,x$. We assume that $A(t)$ is an irreducible Metzler matrix, the flow is strongly monotone.

Proof
Immediately from the preceding theorem. The solution of

$$\left\{\begin{array}{ll} \dot{x} & = A(t)\,Xx \\ x(t_0) & = x_0 \end{array}\right.$$

is given by $x(t, t_0, x_0) = R(t, t_0)\,x_0$. Preceding theorem shows that $R(t, t_0) \gg 0$. Therefore if $x > y$ then $R(t, t_0)\,x \gg R(t, t_0)\,y$. ■
Following Theorem is a sufficient condition for the strong monotonicity

Theorem 5.6.2
Let F a class $\mathcal{C}^1$ vector field on open convex set Ω. If the Jacobian $\frac{\partial F}{\partial x}$ is an irreducible Metzler, the vector field F is strongly monotone.

Proof
We use the relation

$$x(t,t_0,x_1)-x(t,t_0,x_0)=\left(\int_0^1 \frac{\partial x}{\partial x_0}(t,t_0,s\,x_1+(1-s)\,x_0)\,ds\right)\,(x_1-x_0)$$

We denote by $\frac{\partial x}{\partial x_0}(t,t_0,x_0)$ the derivative of the solution of $\dot{x}=F(x)$ with respect to the initial condition.
We know that $\frac{\partial x}{\partial x_0}(t,t_0,x_0)$ is the fundamental solution of the non-autonomous ODE

$$\left\{\begin{array}{ll}\dot{X} & =\frac{\partial F}{\partial x}\left(x(t,t_0,x_0)\right)\,X\\ X(t_0) & =I\end{array}\right.$$

By hypothesis $\frac{\partial F}{\partial x}\left(x(t,t_0,x_0)\right)$ is irreducible Metzler, hence from the corollary (5.6.1) the flow, in other words $\frac{\partial x}{\partial x_0}(t,t_0,x_0)$ is strongly monotone. Since $x_1>x_0$ we deduce the strong inequality in the integral and therefore for the integral, , i.e. $x(t,t_0,x_1)\gg x(t,t_0,x_0)$.

■

5.7 A convergence Criteria

We have the following proposition

Proposition 5.7.1 *: [Hirsch [43]]*
We consider a monotone vector field and an initial condition x, such that the positive trajectory from x is bounded. Assume that there exits $T>0$ such that

$$\phi_T(x)\geq x \quad ou \quad \phi_T(x)\leq x$$

Then the ω-limit set of the trajectory is periodic, with period T.

Proof
Assume $\phi_T(x)\geq x$. If $\phi_T(x)=x$ we are finished. If x is not a fix-point, then $\phi_T(x)$ is different of x, i.e., $x<\phi_t(x)$. Since the trajectory is relatively compact, we can extract a convergent subsequence from the sequence $\phi_{n\,T}(x)$, that we will denote by $\phi_{n_k\,T}(x)$, in such a manner that n_k is strictly increasing and

$$\lim_{k\to\infty}\ \phi_{n_k\,T}(x)=y$$

This limit is a point of the y omega-limit set $\omega(x)$ of the trajectory.
Since the vector fiels is monotone, and since by hypothesis $\phi_T(x)>x$, we deduce by induction, that for any pair of natural number $n>p$ owe have

$$\phi_{n\,T}(x) > \phi_{p\,T}(x)$$

Therfeore

$$\phi_{n_{k+1}\,T}(x) \geq \phi_{(n_k+1)\,T}(x) = \phi_T\left[\phi_{n_k\,T}(x)\right] > \phi_{n_k\,T}(x)$$

going to the limit, from the inequalities we deduce

$$y = \phi_T(y)$$

The ω-limit point y is on a trajectory of period T. It remains to show that the trajectory is exactly the omega-limit set .
Let $z \in \omega(x)$ a strictly increasing sequence t_p such that $\lim_{p\to\infty} \phi_{t_p}(x) = z$. Let for any p be the index $n_{k(p)}$ of the defining sequence of y such that

$$n_{k(p)}\,T \leq t_p < (n_{k(p)} + 1)\,T$$

The sequence of real $t_p - n_{k(p)}\,T$ is abounded sequence. Then a convergent subsequence can be extracted. To simplify we will denote in an identical way. Then $\lim_{p\to\infty}(t_p - n_{k(p)}\,T) = \tau$. Which gives

$$\phi_{t_p}(x) = \phi_{n_{k(p)}\,T}\left[\phi_{t_p - n_{k(p)}\,T}(x)\right] = \phi_{t_p - n_{k(p)}\,T}\left[\phi_{n_{k(p)}\,T}(x)\right]$$

Sequence $n_{k(p)}\,T$ being an extracted sequence of $n_k\,T$ defining y we deduce, passing to the limit that

$$z = \phi_\tau(y)$$

The omega-limit point is therefore on the periodic trajectory from y.

■

We can now state an useful theorem for existence of an equilbrium

Theorem 5.7.1 *[Hirsch [43]]*
We consider a monotone vector field. Let $\{\phi_t(x) \mid t \geq 0\}$ a relatively compact positive trajectory. If there exists a positive $T > 0$ such that

$$\phi_T(x) \gg x \quad ou \quad \phi_T(x) \ll x$$

then $\phi_t(x)$ converge to an equilibrium when t tends to infinity.

If the field is strongly monotone, it is sufficient to have $x < \phi_t(x)$ or $\phi_t(x) < x$ to conclude.

Proof
Assume $\phi_T(x) \gg x$. Other case is similar. Due to this strict inequality and by continuity of the solution of an ODE, it exists ε such that $\phi_s(x) \gg x$ if $s \in [T-\varepsilon, T+\varepsilon]$.
From proposition 5.7.1, we know that the omega-limit set is periodic trajectory of period s for any $s \neq 0$, $s \in]T-\varepsilon, T+\varepsilon[$. It is known that the set of period is an additive sub-semigroup of the semigroup $(\mathbb{R}_+, +)$

This semigroup is dense (we can have a period as small as we want). This semigroup is closed (continuity of the trajectory), hence this $\mathbb{R}$, i.e., 0 is a period. The omega-limite set is reduced to point. Q.E.D. ■

5.8 Looking for invariant sets and equilibria

Proposition 5.8.1

Let F be a class $\mathcal{C}^1$ monotone vector field. We denote by $\prec$ any of the following binary relation $\leq$, $<$ and $\ll$. If $x \prec y$ then $\phi_t(x) \prec \phi_t(y)$.

Proof
If $x \leq y$ we have $\phi_t(x) \leq \phi_t(y)$by definition. Since $\phi_t()$ is diffeomorphism, it is bijection therefore if $x < y$ then $\phi_t(x) \neq \phi_t(y$ and consequently $\phi_t(x) < \phi_t(y$.
If $x \ll y$ since $\phi_t()$ is monotone, the interval

$$[x, y] = \{z \mid x \leq z \leq y\}$$

is send in $[\phi_t(x), \phi_t(y)]$. Since $x \ll y$ the set $[x, y]$ has a non empty interior and therefore $[\phi_t(x), \phi_t(y)]$ also, because $\phi_t()$ is a diffeomorphism.. We deduce $\phi_t(x) \ll \phi_t(y)$.

■

The following lemma has been proved by Selgrade in 1980 [82], this version comes from [86].

Lemma 5.8.1
Let f a monotone field . Let denotes by $\prec$ one of the relation $<$, $\leq$ or $\gg$.
Then the set

$$A_+ = \{x \in \mathbb{R}^n | f(x) \succ 0\}$$

is positively invariant and the flow is nondecreasing on this set.
The set

$$A_- = \{x \in \mathbb{R}^n | f(x) \prec 0\}$$

is positively invariant and the flow is nonincreasing on this set.

Proof
Let $\phi_t(x)$ the flow associated to the vector field.
Let $z(t) = f(\phi_t(x))$. We have $\dot{z}(t) = Df(\phi_t(x)\, f(\phi_t(x)) = Df(z(t))\, z(t)$. This proves that $f(\phi_t(x))$ is the unique solution of the system

$$\begin{cases} \dot{x} & = Df(x)\, x \\ x(0) & = f(x) \end{cases}$$

This system is monotone, then let positively invariant the 3 sets $\mathbb{R}^n_+$, $\mathbb{R}^n_+ \setminus \{0\}$ and the interior of $\mathbb{R}^n_+$, i.e., $\mathbb{R}^n_{+,*}$.
Then if $x \in A_+$, $f(x)$ belongs to one of this three set, say e.g. K. The solution of the linearized system, starting from $f(x)$ stays in K, hence $\phi_(f(x) \in K$. This proves that A_+ is positively invariant.
Then $\phi_t(x)$ is nonincreasing if $x \in A_+$, at least locally, but this is true necessarily by positive invariance for all the trajectory.

■

Proposition 5.8.2
Consider a monotone system in $\mathbb{R}^n$, $\dot{x} = f(x)$ with f being $\mathcal{C}^1$.

Let $a \ll b$ such that $f(a) \geq 0$ and $f(b) \leq 0$, then $[a, b]$ is positively invariant. The trajectories from a and b are converge to equilbria.

If there is an unique equilibrium p in $[a, b]$ then p is GAS in $[a, b]$

Proof
Let $x \in [a, b]$, since $f(a) \geq 0$, by the preceding theorem $\phi_t(a)$ is non decreasing, hence $a \leq \phi_t(a)$ and by monotonicity $\phi_t(a) \leq \phi_t(x)$. This gives $a \leq \phi_t(a) \leq \phi_t(x)$. By the same argument with b, this proves that $[a, b]$ is positively invariant. Now $\phi_t(a)$ is a nondecreasing function, by lemma (5.8.1), which is bounded hence converging to a limit.
If there is an unique equilibrium p then $\phi_t(a)$ and $\phi_t(b)$ converge to p. Hence

$$a \leq \phi_t(a) \leq \phi_t(x) \leq \phi_t(b) \leq b,$$

proves the convergence to p. The stability is obtained by considering the order interval $[\phi_t(a), \phi_t(b)]$ which are positively invariant and as small as we want for t large enough.

■

5.9 Sublinearity, positive invariance and equilibria

Definition 5.9.1 (Hirsch-Smith [49])
A map $T : \mathbb{R}^n_+ \longrightarrow \mathbb{R}^n$ is sublinear if

$$0 < \lambda < 1\,, x > 0 \;\Rightarrow \lambda\, T(x) \leq T(\lambda\, x),$$

strictly sublinear if

$$0 < \lambda < 1, x \gg 0 \Rightarrow \lambda\, T(x) < T(\lambda\, x),$$

and strongly sublinear if

$$0 < \lambda < 1, x \gg 0 \Rightarrow \lambda\, T(x) \ll T(\lambda\, x).$$

Strong sublinearity is the strong concavity assumption of Krasnosel′skiĭ [58].

Proposition 5.9.1
If the application T is $\mathcal{C}^1$ then the condition

$$x \gg y \gg 0 \Longrightarrow DT(x) < DT(y)$$

implies the strict sublinearity if $T(0) \geq 0$.
This is called the strict anti-monotonicity of DT

Proof
Let $x \gg 0$ and $\lambda \in (0,1)$ let $\Phi(s) = T(\lambda\, s\, x) - \lambda\, (T(s\, x))$.
We have $\Phi(1) - \Phi(0) = T(\lambda\, x) - \lambda\, T(x) - (1-\lambda)\, T(0)$ and therefore

$$T(\lambda\, x) - \lambda\, T(x) = (1-\lambda)\, T(0) + \left[\int_0^1 (DT(\lambda\, t\, x) - DT(t\, x))\, dt\right]\, (\lambda\, x) > 0$$

Remark 5.9.1
The condition is sufficient but not necessary. The Ricker function $\alpha\, x\, e^{-\beta\, x}$ is strictly sublinear but its derivative is not anti-monotone.

Proposition 5.9.2 *[49]*
The application $T : \mathbb{R}^n_+ \longrightarrow \mathbb{R}^n$ is strongly sublinear provided T is $\mathcal{C}^1$ and for any $x \gg 0$

$$T(x) \gg DT(x)\,.\,x$$

Proof
Let define $\varphi(s) = \dfrac{1}{s}\, T(s\, x)$ then $\varphi'(s) = -\dfrac{1}{s^2}\, T(s\, x) + \dfrac{1}{s}\, DT(s\, x)\, .\, x.$
Since $\varphi'(s) = -\dfrac{1}{s^2}\, [T(s\, x) - DT(s\, x)\, .\, s\, x]$, then by hypothesis $\varphi'(s) \ll 0$, therefore for $0 < \lambda < 1$

$$T(\lambda\, x) - \lambda\, T(x) = \lambda\, [\varphi(\lambda) - \varphi(1)] = -\lambda \int_{\lambda}^{1} \varphi'(s)\, ds \gg 0$$

■

Proposition 5.9.3
Let F be a $\mathcal{C}^1$ vector field in $\mathbb{R}^n$, whose flow ϕ preserves $\mathbb{R}^n_+$ for $t \geq 0$ monotone and strictly sublinear in $\mathbb{R}^n_+$. Then the flow $\phi_t(\,)$ associated to F is monotone and strictly sublinear.
Moreover if F is strongly monotone and strictly sublinear, then $\phi_t(\,)$ is strongly monotone and strongly sublinear.

Proof
The application ϕ_t is monotone since the vector field is monotone (Theorem (5.6.2)).
To prove the strict sublineariry we consider, for $\lambda \in (0, 1)$ and $x \gg 0$, the quantity

$$y(t) = \phi_t(\lambda\, x) - \lambda\, \phi_t(x).$$

We consider the time derivative $\dot{y}$

$$\begin{aligned} \dot{y} &= F(\phi_t(\lambda\, x)) - \lambda\, F(\phi_t(x)) \\ &= F(\phi_t(\lambda\, x)) - F(\lambda\, \phi_t(x)) + [F(\lambda\, \phi_t(x)) - \lambda\, F(\phi_t(x))] \\ &= A(t)\, y + B(t). \end{aligned}$$

With

$$A(t) = \int_0^1 DF\, (s\, \phi_t(\lambda\, x)) + (1 - s)\, \lambda\, \phi_t(x))\; ds,$$

and

$$B(t) = F(\lambda\, \phi_t(x)) - \lambda\, F(\phi_t(x))$$

In other words $y(t)$ is the solution of a linear equation with a second term $B(t)$, and initial condition $y(0) = 0$. We denote by $R(t, t_0)$ the fundamental solution of

$$\left\{ \begin{array}{rl} \dot{X} & = A(t)\, X \\ X(t_0) & = I_n \end{array} \right.$$

Since $\phi_t(\,)$ preserves the nonnegative orthant we have $\phi_t(0) \geq 0$ and since $x \gg 0$ by monotonicity $0 \leq \phi_t(0) \ll \phi_t(x)$ (proposition (5.8.1)). Hence, since F is strictly sublinear, $B(t) > 0$ for any $t \geq 0$.
Then by the variation of constant formula we have

$$y(t) = \int_0^t R(t,\tau)\, B(\tau)\, d\tau.$$

Since for any $z \geq 0$, $DF(z)$ is Metzler, the same is true for $A(t)$. Hence $R(t, t_0) \geq 0$ for $t \geq t_0$ and is a nonsingular matrix. By hypothesis $B(\tau) > 0$, hence $R(t,\tau)\, B(\tau) > 0$, therefore $y(t) \gg 0$ for $t > 0$. Which proves that ϕ_t is strictly sublinear.
The second assertion comes from the observation that $\phi_t(\,)$ is strongly monotone by Theorem (5.6.2) and by noticing that since $DF((x)$ is Metzler irreducible, then $R(t, t_0$ is positive, which implies that, with $B(t) > 0$, $R(t,\tau)\, B(\tau) \gg 0$. This proves the strong sublinearity.

■

The sublinear applications have nice properties. They will be detailed in the next propositions.

Proposition 5.9.4 (Krasnosel′skiĭ sublinearity trick)
Let F be a $\mathcal{C}^1$ vector field in $\mathbb{R}^n$, whose flow ϕ preserves $\mathbb{R}^n_+$ for $t \geq 0$ and is strongly monotone strictly sublinear in $\mathbb{R}^n_+$. Assume that all trajectories are forward complete.
Then F cannot have two distinct positive equilibria in the interior of the nonnegative orthant $\mathbb{R}^n_+$

Proof
Denote by $\phi_(\,)$ the flow of F. The fact that $\phi_t(\,)$ is strongly monotone and strongly sublinear proves the uniqueness of any positive equilibrium p, by Krasnosel′skiĭ sublinearity trick [42, 58] :
Let p_1 and p_2 two different positive equilibrium. There are fix points of ϕ_t. Let r be defined by

$$r = \min\left\{\frac{1}{\|p_1\|_{p_2}}, \frac{1}{\|p_2\|_{p_1}}\right\},$$

such that $p_1 \geq r\, p_2$ and $p_2 \geq r\, p_1$. We have $r \in (0,1)$ since $\|p_1\|_{p_2}\, \|p_2\|_{p_1} > 1$. Indeed let i_0 and j_0 indices realizing the maximum for each of the two weighted norms, then $\dfrac{p_{1,i_0}}{p_{2,i_0}}\, \dfrac{p_{2,j_0}}{p_{1,j0}} > \dfrac{p_{1,i_0}}{p_{2,i_0}}\, \dfrac{p_{2,i_0}}{p_{1,i0}} = 1$.
Actually r is the maximum number such that simultaneously $p_1 \geq r\, p_2$ and $p_2 \geq r\, p_1$.

By strong sublinearity and monotonicity we have

$$p_1 = \phi_t(p_1) \geq \phi_t(r\,p_2) \gg r\,\phi_t(p_2) = r\,p_2.$$

Similarly we have $p_2 \gg r\,p_1$. This contradicts the maximality of r.

■

Now we will revisit proposition (5.8.2) with the additional hypothesis of sublinearity.

Proposition 5.9.5
Consider a $\mathcal{C}^1$ vector field F strongly monotone and strictly sublinear in the nonnegative orthant.
We assume that there exists $0 \ll a \ll b$ such that $F(a) \geq 0$ and $F(b) \leq 0$.
Then the nonnegative orthant is positively invariant and there is an unique positive equilibrium which is GAS on the interior of the nonnegative orthant.

Proof
We know that $[a, b]$ is positively invariant. But we have, by sublinearity, for $0 < \lambda < 1$ the inequalities $F(\lambda\,a) > \lambda\,F(a) \geq 0$ and $\lambda\,F\left(\frac{1}{\lambda}\,b\right) < F(b) \leq 0$. Hence for any $1 > \varepsilon > 0$ and $\xi > 1$ the order interval $[\varepsilon\,a, \xi\,b]$ is positively invariant. This proves that the interior of the nonnegative orthant is positively invariant and that all the forward trajectories are bounded, hence F is a complete vector field.
By proposition (5.8.2) $\phi_t(a)$ converge to an equilibrium, which is unique by sublinearity. This ends the proof.

■

The following proposition is a result of stability with something less stringent that sublinearity. It is used in some proof in [44, 70]

Proposition 5.9.6
Consider a $\mathcal{C}^1$ vector field f strongly monotone
Assume that $x^* \gg 0$ is an equilibrium such that

$$\textit{for any} \quad \lambda > 1 \quad f(\lambda\,x^*) > 0 \ \textit{and for any} \quad \lambda < 1 \quad f(\lambda\,x^*) < 0$$

Then x^* is GAS in $\mathbb{R}^n_+$.

If f is only monotone if

$$\textit{for any} \quad \lambda > 1 \quad f(\lambda\,x^*) \gg 0 \ \textit{and for any} \quad \lambda < 1 \quad f(\lambda\,x^*) \ll 0$$

then x^* is GAS in the interior of the nonnegative orthant.

Proof
For any $\xi < 1$ and $\lambda > 1$ the order interval $B_{\xi,\lambda} = [\xi\, x^*, \lambda\, x^*]$ is positively invariant with f monotone.

- If f is strongly monotone, if $x \in B_{\xi,\lambda}$ we claim that $f(x)$ points into $B_{\xi,\lambda}$. If x in the boundary of $B_{\xi,\lambda}$ either $\xi\, x^* < x$ or $x < \lambda\, x^*$. Since f is strongly monotone and increasing at $\xi\, x^*$ (respectively decreasing at $\lambda\, x^*$), then either for $t > 0$ we have $\xi\, x^* \leq \phi_t(\xi\, x^*) \ll \phi_t(x)$ or $\phi_t(x) \ll \phi_t(\lambda\, x^*) \leq \lambda\, x^*$, i.e.,

$$\phi_t\left(B_{\xi,\lambda}\right) \subset \overset{\circ}{B}_{\xi,\lambda}$$

- If f is monotone, but since $f(\xi\, x^*) \gg 0$ and $f(\lambda\, x^*) \ll 0$, we have $\xi\, x^* \ll \phi_t(\xi\, x^*) \leq \phi_t(x)$ and $\phi_t(x) \leq \phi_t(\lambda\, x^*) \ll \lambda\, x^*$, again

$$\phi_t\left(B_{\xi,\lambda}\right) \subset \overset{\circ}{B}_{\xi,\lambda}$$

This proves that there is no other equilibrium in $B_{\xi,\lambda}$. Then by proposition (5.8.2), the unique equilibrium x^* is GAS in any $B_{\xi,\lambda}$, hence in the interior of the nonnegative orthant for f monotone and the nonnegative orthant for f strongly monotone.

5.10 A Theorem on stability

The following theorem give a simple proof of Theorem 6.1 proved by Hirsch [44] and also a theorem with the weaker condition given by Smith [85]. The strict antimonotonicity of the Jacobian of the vector field in [44] is replaced by strict sublinearity of the vector field. Strict sublinearity implies, as we have seen, strict antimonotonicity of the Jacobian [58, 59].

Theorem 5.10.1 *[Hirsch 1984]*
Let F be a $\mathcal{C}^1$ vector field in $\mathbb{R}^n$, whose flow ϕ preserves $\mathbb{R}^n_+$ for $t \geq 0$ and is cooperative, irreducible and strictly sublinear in $\mathbb{R}^n_+$. Assume that all trajectories in $\mathbb{R}^n_+$ are bounded.

- *If $F(0) = 0$, the origin is an equilibrium, then either all trajectories in $\mathbb{R}^n_+$ tend to the origin, or else there is a unique equilibrium $p \gg 0$ which is globally asymptotically stable on $\mathbb{R}^n_+ \setminus \{0\}$.*

- *If $F(0) > 0$, there is a unique equilibrium $p \gg 0$ which is globally asymptotically stable on $\mathbb{R}^n_+$.*

Proof
Consider the case that the trajectory of some $x \in \mathbb{R}^n_+$ does not tend to 0. By hypothesis this trajectory is forward bounded. The omega limit set $\omega(x)$ of x has a least upper bound Y in $\mathbb{R}^n_+ \setminus \{0\}$, for the ordering on $\mathbb{R}^n$. Let z be a point of $\omega(x)$. Since $\omega(x)$ is an invariant set, $\phi_{-t}(z) \in \omega(x)$ for any $t \geq 0$. Since ϕ is a monotone flow, from $\phi_{-t}(z) \leq Y$ we deduce $z \leq \phi_t(Y)$ for any $t \geq 0$. This proves that $\phi_t(Y)$ is an upper bound of $\omega(x)$, whence $Y \leq \phi_t(Y)$. If Y is an equilibrium, we have finished, otherwise $Y < \phi_t(Y)$, and by strong monotonicity of the flow we deduce that, for a $T > 0$, we have $Y \ll \phi_T(Y)$. By theorem 5.7.1, since the system is monotone and the trajectories bounded, $\phi_t(Y)$ converges to an equilibrium $p > 0$ as $t \to +\infty$. Moreover by strong monotonicity since $p > 0$ implies $p = \phi_t(p) \gg \phi_t(0) \geq 0$.

This proves that any trajectory, which does not tend to the origin, converges to a positive equilibrium, which is unique by sublinearity (proposition (5.9.4))
By sublinearity, for any $s \in (0,1)$, $F(s\,p) > s\,F(p) = 0$. Again for any $\lambda > 1$ we have $\frac{1}{\lambda} f(\lambda\,p) > f(p) = 0$. We are in the situation of proposition (5.9.6). Hence p is GAS in the interior of the nonnegative orthant. By strong monotonicity, any trajectory, starting from a face of the orthant, except at the origin, enters the positive orthant. Then no trajectory of $\mathbb{R}^n_+ \setminus \{0\}$ tends to zero, hence tends to p which is GAS.

■

Remark 5.10.1
This theorem can extend to any compact subset $K \subset \mathbb{R}^n_+$ positively invariant. Strong monotonicity can be only required in the interior of K as long as the boundary of K is not positively invariant.
Exercise : prove it
See example on Schistosomiasis below.

5.11 Another Theorem from Hirsch

Theorem 5.11.1
Assume that X is a strongly ordered separable topological vector space. Assume ϕ strongly monotone in X (or SOP) and that every orbit of ϕ is closure compact. Let p, q be equilibria with $p \ll q$, with no other equilibria in the order interval $[p,q]$. Then: either every trajectory in $[p,q] \setminus \{p\}$ approaches q, or else every trajectory in $[p,q] \setminus \{q\}$ approaches p.

comments:

This theorem has to be applied on a strongly ordered separable topological vector space.

We say ϕ is order-compact if $\phi_t(S)$ has compact closure whenever $t > 0$ and S in the domain $D(\phi_t()$ of $\phi_t()$ is order-bounded.

$D(\phi_t)$ is the domain of ϕ_t an open subset of X.

The set of equilibria is denoted by E.

A point $x \in X$ is quasiconvergent if $\omega(x) \in E$; the set of quasiconvergent points is denoted by Q.

We call x convergent when $\omega(x)$ is a singleton $\{p\}$; in this case $\phi_t(x) \longrightarrow p \in E$.

The set of convergent points is denoted by C.

When all orbit closures are compact and E is totally disconnected (e.g., countable), then $Q = C$; because in this case every omega limit set, being a connected subset of E, is a singleton.
We call ϕ strongly order preserving, SOP for short, if it is monotone and whenever $x < y$ there exists open subsets U and V of x,y respectively, and $t_0 > 0$ such that

$$\phi_{t_0}(U) \leq \phi_{t_0}(V)$$

Monotonicity of ϕ then implies that for any $t \geq t_0$ $\phi_t(U) \leq \phi_t(V)$.

We say that ϕ is eventually strongly monotone if it is monotone and whenever $x < y$ there exists $t_0 > 0$ such that

$$t \geq t_0 \Longrightarrow \phi_t(x) \ll \phi_t(y)$$

If ϕ is eventually strongly monotone then it is SOP.

5.12 Hirsch's Theorem modified

In the following we denote by $K = \mathbb{R}^n_+$ (some other cone to see ...) the nonnegative orthant.
Recall definition (**??**) : A subset F of K is called a face if F is a cone and if for any $x \in F$

$$0 \leq y \leq x \Longrightarrow y \in F$$

Definition 5.12.1
We consider an order interval $[p, q]$ with $p \ll q$. We have

$$[p,q] = p + K \cap q - K$$

$[p,q]$ is a convex polytope, i.e, the finite intersection of half-spaces. A hyperplane H of R^n is supporting $[p,q]$ if one of the two closed halfspaces of H contains $[p,q]$. A subset F of $[p,q]$ is called a face of $[p,q]$ if it is either $\emptyset$, $[p,q]$ itself or the intersection of $[p,q]$ with a supporting hyperplane.

Theorem 5.12.1
We consider a $\mathcal{C}^1$ monotone system $\dot{x} = f(x)$, whose flow ϕ preserve $\mathbb{R}^n_+$ for $t \geq 0$. We have two equilibria $p \ll q$, with no other equilibrium in $[p,q]$.
We assume that p is asymptotically stable for ϕ and that there exists a positively invariant face F of $[p,q]$ containing p which is in the basin of p. We also assume that some other faces $F_1, F_2, \cdots, F_K$ are positively invariant, and for any point x of each F_i, the omega-limit set satisfies $\omega(x) \cap F \neq \emptyset$.
We assume that the system is strongly monotone on the complement

$$[p,q] \setminus (F \cup F_1 \cup \cdots \cup F_k).$$

Then any trajectory of $[p,q] \setminus \{q\}$ converges to p. Hence the equilibrium q is unstable.

Proof
Consider the totally ordered arc in $[p,q]$, $J = \{x = (1-t)\,p + t\,q \mid 0 \leq t \leq 1\}$. If point $x \in I$ converges to p then by monotonicity $[x,p]$ is in the basin of p. Since p is asymptotically stable there exists $x \neq p$ in J converging to p. Let $a = \sup\{x \in J \mid x \text{ converges to } p\}$. Similarly let $b = \inf\{x \in J \mid x \text{ converges to } q\}$. The order interval $[[a,b]]$ is composed of points converging neither to p, neither to q.
The ordered interval $[p,a]]$ is an open subset of $[p,q]$, containing p in the basin of p.
We will show that $a = b = q$ which will prove the theorem.
Before this we show that any face F_i is in the basin of p. We know that for any x in F_i, $\omega(x) \cap F \neq \emptyset$. Since any point of F converges to p, by invariance of the omega limit set, we have $p \in \omega(x)$, which implies that the trajectory from x enters $[p,a]]$ proving that x converges to p.
We proceed by contradiction assuming $a \ll b$. By definition of a, any point of $[[a,b]]$ does not converge to p. Let

$$W = \overline{\phi_{\mathbb{R}_+}([[a,b]] \cup J)}.$$

Continuity of ϕ implies that W is a separable metric space positively invariant under ϕ. Moreover $W \cap (F \cup F_1 \cup \cdots \cup F_k) = \emptyset$, otherwise this will implies that

a point of $[[a,b]]$ converges to p, a contradiction. It follows that W is an ordered space, with a strongly monotone semiflow ϕ. By Theorem 1.14 of [49], if Q is the set of quasiconvergent points, $([[a,b]] \cup J) \setminus Q$ is at most countable. Then there is a point in $[[a,b]] \cup J$ converging to p or to q. A contradiction, hence $a = b$.
We will now prove that $a = q$. We proceed by contradiction, assuming that $a \ll q$. We already know that $p \ll a$. The point a cannot converge to p. If not, since the basin of an asymptotically stable point is open, this will contradict the definition of a. If a converges to q, this will implies that in $[p,q]$ the point q is asymptotically stable, by the preceding argument we obtain again a contradiction. This proves that a is a nonconvergent point. The set $X = \overline{\phi_{\mathbb{R}_+}(a)}$ is a positively invariant compact ordered set with a strongly monotone semiflow ϕ without equilibrium. By lemma 1.1 of [49], X has a maximal element z. For any $t \geq 0$, $\phi_t(z)$ is an upper bound of X, hence $z = \phi_t(z)$ is an equilibrium, a contradiction. This ends the proof of the Theorem.

■

5.13 Example : Gonorrhea

Gonorrhea is a sexually transmitted disease caused by the bacterium *Neisseria gonorrhoeae* manifested primarily by urethritis in men, vaginitis, cervicitis and metritis in women.
Neisseria gonorrhoeae is a strict human parasite, host of the mucous membranes of the genital tract of man and woman. Transmission is mainly direct (this germ being fragile) and almost venereal.
This is one of the most common infectious diseases with more than 200 million annual cases worldwide. It primarily affects the poor.
A person treated may again be contaminated. There 's no immunization for gonorrhea.
The following model is a seminal paper from Lajmanovitch and Yorke [60] in 1976. In 1984 Hirsch remarks the monotonicity of the system [44] and revisits stability analysis.
We consider n groups (patches). We begin with a model in one isolated group : SIS model

$$\left\{ \begin{array}{rl} \dot{S} & = \Lambda - \mu\, S - \beta\, \frac{S}{N}\, I + \gamma\, I \\ \dot{I} & = \beta\, \frac{S}{N}\, I - (\gamma + \mu)\, I \end{array} \right. \tag{5.3}$$

N is the total population

$$\left\{ \begin{array}{rl} \dot{N} & = \Lambda - \mu\, N \\ \dot{I} & = \beta\, \frac{S}{N}\, I - (\gamma + \mu)\, I \end{array} \right. \tag{5.4}$$

This system is triangular, $N^* = \frac{\Lambda}{\mu}$ and we can apply Vidyasagar's Theorem (3.6.1). It is sufficient to study the reduced system :

$$\dot{I} = \frac{\beta}{N^*}\,(N^* - I)\,I - (\gamma + \mu)\,I$$

$$x = \frac{I}{N^*} \quad \dot{x} = \tilde{\beta}\,x(1-x) - \tilde{\gamma}\,x$$

Model in n groups : multigroup model. n equations $i = 1, \ldots, n$.

$$\dot{x}_i = (1 - x_i) \sum_j \beta_{i,\,j}\,x_j - \alpha_i\,x_i \tag{5.5}$$

$\beta_{i,\,j}\,x_j$ visiting infectious from patch j coming in patch i

$$\dot{x} = \Big[D + B - \mathrm{diag}(x)\,B\Big]\,x$$

$B = (\beta_{i,j})$,
$D = -\mathrm{diag}(\alpha_i)$.

$$\dot{x} = \Big[D + B - \mathrm{diag}(x)\,B\Big]\,x$$

rewritten in Scilab/Matlab notations

$$\dot{x} = \Big[D + \mathrm{diag}(1-x)\,B\Big]\,x$$

$$f(x) = \Big[D + \mathrm{diag}(1-x)\,B\Big]\,x$$

$$\mathrm{Jac} f(x) = D + \mathrm{diag}(1-x)\,B - \mathrm{diag}(Bx)$$

B irreducible, system is strongly monotone and strictly sublinear.
Remark

$$B + D \text{ stable} \Longleftrightarrow s(B+D) < 0 \Longleftrightarrow \mathcal{R}_0 = \rho(-B\,D^{-1}) > 1$$

Varga : $B + D$ regular splitting $B \geq 0$ D stable Metzler matrix.
Diekmann, van den Driessche-Watmough : $-B\,D^{-1}$ next generation matrix.
All the condition of Hirsch's Theorem are satisfied hence

Theorem 5.13.1
If $\rho(-B^{-1}) < 1$ the origin (DFE) is GAS for system (5.5).
If $\rho(-B^{-1}) > 1$ there exists a unique positive endemic equilibrium, which is GAS for system (5.5) on $[0,1] \setminus \{0\}$.

Remark 5.13.1
What happens when $\rho(-B^{-1}) = 1$? It can be shown that the DFE is GAS using Lyapunov techniques [30].

5.14 Ross model in a patchy environment

5.14.1 The migration model

This model does not keep track of where an individual usually resides, but just considers where he is at time t. As in the Ross model, the demography is neglected, i.e., there is no death or birth. The transfer rate from patch i to patch j, for $i \neq j$, is denoted by $m_{ji} \geq 0$. The total host population on patch i is denoted N_i. Hence, for $i = 1, \cdots, n$, the dynamics is given by

$$\dot{N}_i = \sum_{j=1\, j\neq i}^{n} m_{ij}\, N_j - N_i \sum_{j=1\, j\neq i}^{n} m_{ji}\ .$$

This system can be written

$$\dot{N} = M\, N\ . \tag{5.6}$$

Where N is the column vector $(N_1, \cdots, N_n)^T$, the superscript T denotes transpose, and the matrix M is defined by $M(i,j) = m_{ij}$, for $i \neq j$ and

$$M(i,i) = -\sum_{j=1\, j\neq i}^{n} m_{ji}\ .$$

5.14.2 The Ross-Macdonald model on n patches

We use the following notations :

- $I_{h,i}$ is the infectious host population on patch i.
- p is the number of patches harboring vectors. $I_{v,i}$, V_i are respectively the infectious vector population and the constant vector population on patch i. If $i > p$ there is no vector on patch i, i.e., $V_i = 0$.
- a is the man biting rate of vectors.
- b_1 is the proportion of infectious bites on hosts that produce a patent infection.

- b_2 is the proportion of bites by susceptible vectors on infectious hosts that produce a patent infection.
- μ is the per capita rate of vector mortality.
- γ is the per capita rate of host recovery from infection.

As in the classical Ross-MacDonald model the total population of hosts is constant when we consider all the patches. On the other hand since there are no migration of vectors, the population of vectors is constant on each patch.

Remark 5.14.1 *For sake of simplicity, we assume that the parameters b_1, b_2, γ and μ are the same for all patches. However, the analysis presented here can be extended when these parameters differ from patch to patch.*

We number the patches in such a way that only the p first patches, $1 \leq p \leq n$, are infested by vectors. On the patches i for $i > p$, we have $V_i = 0$ hence, $I_{v,i} = 0$.
Since we already know the dynamics of the host population on patches by system (5.6), it is sufficient to model the dynamics of the population of infectious hosts. Similarly the population of vectors being constant on each patch we only need to study the dynamics of the population of infectious vectors.
For patches such that $i \leq p$, i.e., where vectors are present, we have

$$\begin{cases} \dot{I}_{h,i} = & b_1\, a\, I_{v,i}\, \dfrac{N_i - I_{h,i}}{N_i} - \gamma\, I_{h,i} + \displaystyle\sum_{j=1, j\neq i}^{n} m_{ij}\, I_{h,j} - I_{h,i} \left(\sum_{j=1, j\neq i}^{n} m_{ji} \right) \\ \\ \dot{I}_{v,i} = & b_2\, a\, (V_i - I_{v,i})\, \dfrac{I_{h,i}}{N_i} - \mu\, I_{v,i}\ . \end{cases} \tag{5.7}$$

In the equations for infectious hosts, the term $b_1\, a\, I_{v,i}\, \frac{N_i - I_{h,i}}{N_i}$ corresponds to the infection of susceptible hosts bitten by infectious vectors, using the classical frequency dependent transmission, with a varying host population on patch i. The term $-\gamma\, I_{h,i}$ is the recovery term. The other terms account for migration. In the equations for infectious vectors, $b_2\, a\, (V_i - I_{v,i})\, \frac{I_{h,i}}{N_i}$ corresponds to the infection of susceptible vectors, when biting an infected host. The last term $-\mu\, I_{v,i}$ corresponds to mortality.
For $i > p$, there are no vector on patch i and the equations for infectious hosts only have recovery and migration terms. The equation governing the evolution of $I_{h,i}$ is the following

$$\dot{I}_{h,i} = \quad -\gamma\, I_{h,i} + \sum_{j=1, j\neq i}^{n} m_{ij}\, I_{h,j} - I_{h,i} \left(\sum_{j=1, j\neq i}^{n} m_{ji} \right) . \tag{5.8}$$

The complete system of $2n + p$ equations is obtained by incorporating to the previous system, the n equations $\dot{N} = M\, N$.
We are going to "vectorialize" the model, using the following vectors of $\mathbb{R}^n$:
$I_h = (I_{h,1}, \cdots, I_{h,n})^T$, $I_v = (I_{v,1}, \cdots, I_{v,p}, 0, \cdots, 0)^T$ and
$V = (V_1, \ldots, V_p, 0, \cdots, 0) \in \mathbb{R}^n$.
We denote $\beta_1 = b_1\, a$ and $\beta_2 = b_2\, a$. If X is a vector, with either $X \in \mathbb{R}^p$ or $X \in \mathbb{R}^q$, we denote by $\mathrm{diag}(X, p, q)$ the $p \times q$ matrix whose diagonal is given by the components of X and the other terms are zero. The short notation $\mathrm{diag}(X)$ denotes, if $X \in \mathbb{R}^p$, the $p \times p$ diagonal matrix $\mathrm{diag}(X, p, p)$.

We use the notation π to denote the projection of $\mathbb{R}^n$ on $\mathbb{R}^p$, $p \leq n$;

$$\pi : (x_1, \ldots, x_n)^T \longmapsto (x_1, \ldots, x_p)^T$$

With these notations and conventions, the complete system becomes

$$\begin{cases} \dot{N} = & M\, N \\ \\ \dot{I}_h = & \beta_1\, \mathrm{diag}(N)^{-1}\, \mathrm{diag}(N - I_h)\, I_v - \gamma\, I_h + M\, I_h \\ \\ \pi \dot{I}_v = & \beta_2\, \mathrm{diag}(\pi\, N)^{-1}\, \mathrm{diag}(\pi(V - I_v))\, \pi I_h - \mu\, \pi I_v\ . \end{cases} \tag{5.9}$$

This system evolves on the affine hyperplane of $\mathbb{R}^{2n+p}$, whose equation is $\sum_i N_i = H$, where H is the total host population.

Remark 5.14.2 *In the case where the parameters β_1, β_2, γ and μ are not the same for all patches, they are replaced in system (5.9) by diagonal nonnegative matrices and this does not change the fundamental structure of the system.*

5.14.3 Properties of the model

The model is such that the entire population $H = N_1 + \cdots + N_n$ is constant.
Let $\mathbb{1}$ be the vector $(1, \cdots, 1)^T$ of $\mathbb{R}^n$. The vector $(H, \cdots, H)$ will be denoted $H\, \mathbb{1}$.

Proposition 5.14.1 *:*
The Parallelepiped

$$\mathcal{P} = \{(N, I_h, \pi\, I_v) \in \mathbb{R}^{2n+p} |\, 0 \leq N \leq H\, \mathbb{1};\ 0 \leq I_h \leq H\, \mathbb{1}\ ;\ 0 \leq I_v \leq V\} .$$

is positively invariant for system (5.9).

Proof

It is sufficient to consider the system on the faces of $\mathcal{P}$ and to show that on each face, the vector fields associated to the system points into the nonnegative orthant.

If $I_{h,i} = 0$ then

$$\dot{I}_{h,i} = b_1\, a\, I_{v,i}\, \frac{N_i}{N_i} + \sum_{j=1, j\neq i}^{n} m_{ij}\, I_{h,j} \geq 0.$$

If $I_{h,i} = H$ then for all $j \neq i$ we have $I_{h,j} = 0$, since the entire population is H and

$$\dot{I}_{h,i} = -\gamma\, H - H \left(\sum_{j=1, j\neq i}^{n} m_{ji} \right) < 0.$$

If $I_{v,i} = 0$, then

$$\dot{I}_{v,i} = b_2\, a\,(V_i)\, \frac{I_{h,i}}{N_i} \geq 0.$$

If $I_{v,i} = V_i$ then

$$\dot{I}_{v,i} = -\mu\, V_i < 0.$$

Finally, since M is a Metzler matrix [53], the nonnegative orthant is positively invariant by the system $\dot{N} = M\, N$.

■

5.14.4 Reduction of the system

We will reduce the stability analysis of (5.9), to the study of a smaller and simpler system.

Let us show that matrix M can be assumed to be irreducible. In other words, we can assume that the graph of the patches is strongly connected. If M is not irreducible, then by renumbering the patches, it can be given the following block triangular structure

$$M = \begin{bmatrix} M_{11} & 0 \cdots & 0 & \\ M_{21} & M_{22} & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ M_{k1} & M_{k2} & \cdots & M_{kk} \end{bmatrix},$$

where the diagonal blocks are irreducible.

Let us consider the blocks M_{ii} corresponding to traps, i.e., groups of compartments for which there are no transfers to the environment. It is easily seen that asymptotically, all the material will be transfered into the traps of the system. From this moment the traps will behave as irreducible groups. For studying the asymptotic properties of the system, it is therefore sufficient to restrain the study to traps. This shows that M can be assumed to be irreducible.

Theorem of Vidyasagar (3.6.1) will permit us to reduce the stability analysis to a smaller system

If we prove that the system $\dot{N} = M\,N$ has an equilibrium, which is globally asymptotically stable, then we can apply theorem 3.6.1. Recall that the stability modulus $s(M)$ of a matrix M is the largest real part of the elements of the spectrum $\mathrm{Spec}(M)$ of M.

$$s(M) = \max_{\lambda \in \mathrm{Spec}(M)} \mathrm{Re}(\lambda).$$

By Perron-Frobenius Theorem (5.4.4) if M is an irreducible Metzler matrix, then there exists a positive eigenvector $\mathbf{w} \gg 0$ of M such that $M\,\mathbf{w} = s(M)\,\mathbf{w}$, and any positive eigenvector is a multiple of $\mathbf{w}$. Moreover the multiplicity of the eigenvalue $s(M)$ is 1.

The Metzler matrix M satisfies $\mathbb{1}^T M = 0$ or equivalently $M^T\,\mathbb{1} = 0$. This implies that $s(M) = 0$ and all the other eigenvalues are with negative real part. An immediate consequence of the relation $\mathbb{1}^T M = 0$ is that any trajectory of the system $\dot{N} = M\,N$ remains in the affine hyperplane orthogonal to vector $\mathbb{1}$ and containing the initial condition $N(0)$.

From the preceding remark on irreducible Metzler matrices, there exists $\mathbf{w} \gg 0$ such that $M\,\mathbf{w} = 0$. Hereafter we denote by $\mathbf{w}$ the unique vector $\mathbf{w} \gg 0$, defined by

$$M\,\mathbf{w} = 0 \text{ and } \sum_{i=1}^{n} \mathbf{w}_i = 1. \tag{5.10}$$

Hence $H\,\mathbf{w} = (H\mathbf{w}_1, \cdots, H\mathbf{w}_n)^T$ is in the hyperplane orthogonal to $\mathbb{1}$, and containing $N(0)$. It is the unique equilibrium of the system in this hyperplane. Since all the nonzero eigenvalues of M have a negative real part, this equilibrium is globally asymptoticaly stable on the hyperplane.

Hereafter we denote by $\bar{N} = H\,\mathbf{w}$ this equilibrium. By application of Vidyasagar's theorem the stability analysis of (5.9) is now reduced to the stability analysis of the system

$$\begin{cases} \dot{I}_h = & \beta_1 \operatorname{diag}(\bar{N})^{-1} \operatorname{diag}(\bar{N} - I_h)\, I_v - \gamma\, I_h + M\, I_h \\ \\ \pi\, \dot{I}_v = & \beta_2 \operatorname{diag}(\pi\, \bar{N})^{-1} \operatorname{diag}(\pi\, (V - I_v))\, \pi\, I_h - \mu\ \pi\, I_v. \end{cases} \tag{5.11}$$

The dimension of this reduced system is $n + p$.
Since the populations in the patches are constant, we will now rewrite the system using the prevalence variables $x_i = I_{h,i}/\bar{N}_i$, $y_i = I_{v,i}/\bar{V}_i$ and the vectorial density on each patch denoted by $m_i = V_i/\bar{N}_i$. Clearly the vectorial densities on the last $n - p$ patches are 0. Accordingly the vectors x, y and $\mathbf{m}$ are in $\mathbb{R}^n$. With these notations the system (5.11) can be rewritten as follows

$$\begin{cases} \dot{x} = & \beta_1 \operatorname{diag}(\mathbf{m}) \operatorname{diag}(\mathbb{1} - x)\, y - \gamma\, x + D\, x \\ \\ \pi\, \dot{y} = & \beta_2\ \operatorname{diag}(\mathbb{1} - \pi\, y)\, \pi\, x - \mu\, \pi\, y. \end{cases} \tag{5.12}$$

The matrix D is defined by $D(i,j) = \dfrac{1}{\overline{N}_i}\, M(i,j)\, \overline{N}_j$. In other words

$$D = \operatorname{diag}(\bar{N})^{-1}\, M \operatorname{diag}(\bar{N}).$$

Since $M\mathbf{w} = M\, \bar{N} = 0$, it follows that

$$D.\mathbb{1} = \operatorname{diag}(\bar{N})^{-1}\, M \operatorname{diag}(\bar{N}).\mathbb{1} = 0.$$

In other words, if we denote by $d_{ij} \geq 0$ the (i,j) entry of D for $i \neq j$ and $-d_{ii} \leq 0$ the (i,i) entry of D, then

$$-d_{ii} + \sum_{j \neq i}^{n} d_{ij} = 0.$$

In this form system (5.12) is clearly a generalization on n patches of the classical Ross-Macdonald model, with Dx as the migration term.
We have the straightforward property

Proposition 5.14.2 *:*
The unit cube $[0,1]^{n+p}$ *is positively invariant for system (5.12).*

We will simplify this system one step further. Hereafter $X = (x, \pi\, y) \in \mathbb{R}^{n+p}$, where I_n is the $n \times n$ identity matrix and we set the block matrices

$$B = \begin{bmatrix} 0 & 0 & I_p \\ 0 & I_{n-p} & 0 \\ I_p & 0 & 0 \end{bmatrix},$$

$$\Delta = \begin{bmatrix} D - \gamma\, I_n & 0 \\ 0 & -\mu\, I_p \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \beta_1\, \mathrm{diag}(\mathbf{m}) & 0 \\ 0 & \beta_2\, I_p \end{bmatrix}.$$

B is nonnegative, Λ is a nonnegative diagonal matrix and Δ is a stable Metzler matrix.

Remark 5.14.3 *These properties are unchanged if we consider different parameters β_1, β_2, γ and μ in each patch.*

We now have

$$\dot{X} = \Lambda\, \mathrm{diag}(\mathbb{1} - X)\, B\, X + \Delta\, X = [\Lambda\, \mathrm{diag}(\mathbb{1} - X)\, B\ + \Delta]\, X. \tag{5.13}$$

We consider the evolution of this system on the unit cube of $\mathbb{R}^{n+p}$.

Proposition 5.14.3 *: The system (5.13) is cooperative and strongly monotone on the unit cube $[0,1]^{n+p}$*

Proof
The Jacobian J of the system (5.13) in $\mathbb{R}^{n+p}$ is

$$J(X) = \Lambda\, \mathrm{diag}(\mathbb{1} - X)\, B + \Delta - \Lambda\, \mathrm{diag}(B\, X).$$

Clearly J is a Metzler matrix, for $0 \leq X \leq \mathbb{1}$, and the system is cooperative on the unit cube.
We will first prove that the Jacobian is an irreducible matrix on $(0,1)^{n+p}$. This will induce the strong monotonicity on the interior of the unit cube. We need to examine more closely the Jacobian. Using the notation $\mathrm{diag}(X, p, q)$, introduced in section 5.14.2, we can decompose $J(X)$ into $n \times n$, $n \times p$, $p \times n$ and $p \times p$ blocks as follows :

$$J(x, \pi\, y) = \begin{bmatrix} D - \gamma\, I_n - \beta_1\, \mathrm{diag}(\mathbf{m})\, \mathrm{diag}(y) & \beta_1 \mathrm{diag}(\pi\, \mathbf{m}, n, p)\, \mathrm{diag}(\mathbf{1} - \pi\, x) \\ \\ \beta_2\, \mathrm{diag}(\mathbf{1} - \pi\, y, p, n) & -\mu\, I_p - \beta_2\, \mathrm{diag}(\pi\, x) \end{bmatrix}.$$

To prove the irreducibility, it is sufficient to check that the directed graph associated with J is strongly connected. For the sake of intelligibility we also name x_i,

$i = 1, \cdots, n$ and y_j, $j = 1, \cdots, p$ the vertices of the associated graph. The irreducibility of D and the structure of J imply that the subgraph generated by the vertices x_i is strongly connected. Now, if $y \ll 1$, the $p \times n$ block $\beta_2 \operatorname{diag}(\mathbf{1} - \pi\, y, p, n)$ shows that there is a path which, from any vertex y_j where $j \leq p$, leads to vertex x_i. In the same manner if $x \ll 1$ there is a path from any vertex x_i to the vertex y_j (necessarily $j \leq p$). It is now clear that between any couple of vertices there exists a path. This proves the strong monotonicity on the unit cube except on the faces, i.e., for $x \not\ll 1$ and $y \not\ll 1$. Now if an initial point is on one of these faces, the trajectory of the system leaves immediately the face. This implies the strong monotonicity of the flow of the system.

■

5.14.5 Main theorem

In this section we will give an analytic expression for $\mathcal{R}_0$ and completely answer to the stability question.
As usual $\rho(M)$ is the spectral radius of a matrix M. To express the basic reproduction ratio, we need a notation to extract blocks from matrices. If M is a matrix, we denote by $M(1:p, 1:q)$ the submatrix consisting of p first rows and the q first columns of M. With this notation

Proposition 5.14.4 *:*
The origin is the DFE of (5.12) and

$$\mathcal{R}_0^2 = \frac{\beta_1\, \beta_2}{\mu}\, \rho\left(-\operatorname{diag}(\pi\, \mathbf{m})\, Z\right),$$

where $Z = (D - \gamma\, I)^{-1}\,(1:p, 1:p)$*, i.e., the submatrix of the p first rows and p first columns of* $(D - \gamma\, I)^{-1}$*.*

Proof
The Jacobian, computed at the DFE, is $J(0) = \Lambda\, B + \Delta$. The part coming from infection is $\Lambda\, B$ and the part coming from other transfers is Δ. Hence $\mathcal{R}_0 = \rho(-\Lambda\, B\, \Delta^{-1})$. If $I_{p,n}$ is the $p \times n$ identity matrix, we set

$$F = \begin{bmatrix} 0 & \beta_1 \operatorname{diag}(\pi\, \mathbf{m}, n, p) \\ \beta_2\, I_{p,n} & 0 \end{bmatrix}.$$

$$V = \begin{bmatrix} D - \gamma\, I_n & 0 \\ 0 & -\mu\, I_p \end{bmatrix}.$$

We have $\mathcal{R}_0 = \rho(-F\,V^{-1}) = \rho(-V^{-1}\,F)$. We can compute $F\,V^{-1}$. We remark that the product $I_{p,n}\,.\,(D-\gamma\,I_n)^{-1}$ is the block (p,n) extracted from the matrix $(D-\gamma\,I_n)^{-1}$.

$$-F\,V^1 = \begin{bmatrix} 0 & \dfrac{\beta_1}{\mu}\,\mathrm{diag}(\pi\,\mathbf{m},n,p) \\ -\beta_2\,(D-\gamma\,n)^{-1}(1:p,1:n) & 0 \end{bmatrix}.$$

If A and B are two $n\times p$ and $p\times n$ matrices, we have the relation

$$\det\begin{bmatrix}\lambda\,I_n & -A\\ -B & \lambda\,I_p\end{bmatrix} = \lambda^{n-p}\det(\lambda^2\,I_p - B\,A) = \lambda^{p-n}\det(\lambda^2\,I_p - A\,B)$$

The last relation implying simply that, in the characteristic polynomial of AB, there are $(n-p)$ roots equal to zero. Then, using again the structure of $\mathrm{diag}(\mathbf{m},n,p)$ and Cayley reduction, we have

$$\mathcal{R}_0^2 = \rho\left(-\frac{\beta_1\,\beta_2}{\mu}\,\mathrm{diag}(\pi\,\mathbf{m})\ \left[(D-\gamma\,I_n)^{-1}(1:p,1:p)\right]\right),$$

where $(D-\gamma\,I)^{-1}(1:p,1:p)$ is the block of the p first rows and p first columns of $(D-\gamma\,I)^{-1}$.

For the classical Ross-Macdonald model, this formula gives the intended result $\mathcal{R}_0^2 = \dfrac{m\,a^2\,b_1\,b_2}{\gamma\,\mu}$. We have now reduced the study of the stability of the complete system (5.9) to the study the stability of (5.13).

Remark 5.14.4 *The expression of $\mathcal{R}_0$ can be rendered more geometrical if we use the projection π. This is useful if the reordering of coordinates is not done.*

$$\mathcal{R}_0^2 = \rho\left(-\frac{\beta_1\,\beta_2}{\mu}\ diag(\pi\,\mathbf{m})\,\pi\,(D-\gamma\,I)^{-1}\,\pi^T\right).$$

Finally, using the definition of D and the fact that diagonal matrices commute, we have the following expression of $\mathcal{R}_0$ for the original system (5.9)

$$\mathcal{R}_0^2 = \frac{\beta_1\,\beta_2}{\mu}\,\rho\left(-\,diag(\pi\,V)\,diag(\pi\,\bar{N})^{-2}\,\pi\,(M-\gamma\,I)^{-1}\,\pi^T\,diag(\pi\,\bar{N})\right). \qquad (5.14)$$

We know from [26, 94] that if $\mathcal{R}_0 < 1$ then the DFE is locally asymptotically stable, and if $\mathcal{R}_0 > 1$ then the DFE is unstable. Our main result is a global stability result of the DFE , which holds for $\mathcal{R}_0 \le 1$ and a global stability result for $\mathcal{R}_0 > 1$.

Theorem 5.14.1 *:*
If $\mathcal{R}_0 \leq 1$, then all the trajectories of (5.13) tend to the disease free equilibrium, which is globally asymptotically stable on the unit cube. If $\mathcal{R}_0 > 1$, then there exists a unique endemic equilibrium $(\bar{x}, \bar{y}) \gg 0$, and all the trajectories of the unit cube, minus the origin, tend to this equilibrium which is GAS on the unit cube minus the origin.

Proof
We recall the system (5.13)

$$\dot{X} = \Lambda \operatorname{diag}(1 - X)\, B\, X + \Delta\, X = [\Lambda \operatorname{diag}(1 - X)\, B \ + \Delta]\, X.$$

To prove the first assertion we assume $\mathcal{R}_0 = \rho(-\Lambda\, B\, \Delta^{-1}) \leq 1$.
The Jacobian at the origin is $J = \Lambda\, B \ + \Delta$. Since Δ is a nonsingular Metzler matrix and $\Lambda \geq 0$, this expression is a regular splitting of J. Then we have seen that we have the equivalence between $s(J) \leq 0$ and $\rho(-\Lambda\, B\, \Delta^{-1}) \leq 1$.
We have seen that the Jacobian at the origin is an irreducible Metzler matrix and that there exists a positive vector $c \gg 0$ such that

$$c^T\, (\Lambda\, B \ + \Delta) = s(J)\, c^T \leq 0.$$

We use on the unit cube the Liapunov proper function

$$V(X) = \langle c\,|X\rangle,$$

where $\langle \ | \ \rangle$ denotes the usual inner product. Since $c \gg 0$ this function is positive definite in the nonnegative orthant. We compute the derivative $\dot{V}$ of V along the trajectories of (5.13)

$$\begin{aligned} \dot{V} \ &= \langle c\,|(\Lambda \operatorname{diag}(1 - X)\, B \ + \Delta)\, X\rangle \leq \langle c\,|(\Lambda\, B \ + \Delta)\, X\rangle = \langle (\Lambda\, B \ + \Delta)^T\, c\,|\, X\rangle \\ &= \langle s(J)\, c|X\rangle = s(J)\, \langle c|X\rangle \leq 0 \end{aligned}$$

This proves the stability of the DFE. We now consider the asymptotic stability.

$$\dot{V}(X) = \langle c\,|(\Lambda \operatorname{diag}(1 - X)\, B \ + \Delta)\, X\rangle = \langle c\,|(\Lambda\, B \ + \Delta)\, X\rangle - \langle c\,|\Lambda \operatorname{diag}(X)\, B\, X\rangle.$$

If $\mathcal{R}_0 < 1$, then $s(\Lambda\, B \ + \Delta) < 0$ whence

$$\dot{V}(X) \leq \langle c\,|(\Lambda\, B \ + \Delta)\, X\rangle = s(\Lambda\, B \ + \Delta)\, \langle c\,|X\rangle < \langle c|X\rangle < 0.$$

When $\mathcal{R}_0 = 1$, we have $\alpha(\Lambda\, B \ + \Delta) = 0$, and $\dot{V}$ reduces to

$$\dot{V}(X) = -\langle c \,|\Lambda \,\mathrm{diag}(X)\, B\, X\rangle \leq 0.$$

We consider the set $\mathcal{E} = \{X \in [0,1]^{n+p} |\ \dot{V}(X) = 0\}$. Since $c \gg 0$ and $\Lambda \,\mathrm{diag}(X)\, B\, X \geq 0$, this set is composed of points for which $x_i\, y_i = 0$, for all the indexes $i \leq p$. We show that the largest invariant set $\mathcal{L}$, contained in $\mathcal{E}$ is reduced to $\{0\}$. The proof hinges on the irreducibility of the system. There exists at least a point in $\mathcal{L}$ such that $x_i = 0$ or $y_i = 0$ with $i \leq p$.

If $x_i = 0$ considering

$$\begin{aligned}\dot{x}_i &= \beta_1\, m_i\, (1 - x_i)\, y_i - \gamma\, x_i + (M\, x)_i \\ &= \beta_1\, m_i\, y_i + \sum_{j \neq i} m_{ij}\, x_j = 0.\end{aligned}$$

This shows that $y_i = 0$, and all the x_j for which $m_{ij} \neq 0$ i.e., all the x_j "connected" to x_i are equal to zero. A finite recursion argument, with the irreducibility hypothesis, concludes that all the x_i and the y_i are equal to zero. The set $\mathcal{L}$ is reduced to the origin.

If $y_i = 0$ considering

$$\dot{y}_i = \beta_2\, (1 - y_i)\, x_i - \mu\, y_i = \beta_2\, x_i = 0.$$

the invariance of $\mathcal{L}$ shows that $x_i = 0$, and we are back to the preceding situation. In any case $\mathcal{L}$ is reduced to the origin. By LaSalle's invariance principle [61] we deduce the global asymptotic stability of the DFE.

which proves the asymptotic stability of the DFE.

For the second assertion of the Theorem, when $\mathcal{R}_0 > 1$ then the DFE is unstable. It is straightforward to check that the system is strongly sublinear. Let $0 < \lambda < 1$ and

$$T(X) = \Lambda \,\mathrm{diag}(1 - X)\, B\, X + \Delta\, X$$

Then if $X \gg 0$, $B\, X \gg 0$

$$T(\lambda\, X) = \Lambda \,\mathrm{diag}(1 - \lambda\, X)\, B\, \lambda\, X + \Delta\, \lambda\, X \gg \Lambda \,\mathrm{diag}(1 - X)\, B\, \lambda\, X + \Delta\, \lambda\, X = \lambda\, T(X).$$

This proves the strong sublinearity, hence all the hypothesis of Hirsch's Theorem (5.10.1) are satisfied, therefore there exists an unique positive equilibrium which is GAS on $[0,1]^{n+p} \setminus \{0\}$.

■

5.15 Wolbachia

Wolbachia is a bacteria which infects arthropod species, including a high proportion of insects (60% of species).

The unique biology of *Wolbachia* has attracted a growing number of researchers.

While *Wolbachia* is commonly found in many mosquitoes it is absent from the species that are considered to be of major importance for the transmission of human pathogens.

The successful introduction of a life-shortening strain of *Wolbachia* into the dengue vector *Aedes aegypti* that decreases adult mean life has recently been reported.

Moreover it is estimated that the population of mosquitoes harboring *Wolbachia* is less efficient to transmit dengue, as some results has been obtained.

Then it is considered that using *Wolbachia* can be a viable option for controlling the incidence of the dengue.

The bacteria is transmitted only by the eggs laid by female mosquitoes.

Our model take into account cytoplasmic incompatibility, which is outlined in the following table :

Table 5.1: Cytoplasmic incompatibility

Reproduction			
		♂	
		Infected	Uninfected
♀	Infected	Infected	Infected
	Uninfected	Sterile	Uninfected

This phenomenon causes embryos from Wolbachia-uninfected females to die when they are mated with infected males whereas infected females are not affected in this manner.

We will give a simple model introduced by Moacyr : L_u is the aquatic stage (eggs, Larvae, puppae) uninfected by *Wolbachia*, L_w infected, A_u and A_w the respectively adult stages uninfected and infected.

There are a intra-specific competition in the aquatic stages and cytoplasmic incompatibility is modeled in equation 2 and 4.

$$\begin{cases} \dot{A}_u &= \lambda\, L_u - \mu_u\, A_u \\ \dot{L}_u &= r\, A_u\, \dfrac{A_u}{A_u + A_w} - [\lambda + \nu + d\,(L_u + L_w)]\, L_u \\ \dot{A}_w &= \lambda\, L_w - \mu_i\, A_w \\ \dot{L}_w &= r\, A_w - [\lambda + \nu + d\,(L_u + L_w)]\, L_w \end{cases} \tag{5.15}$$

This system is defined on the nonnegative orthant, by Lipschitz prolongation. The nonnegative orthant is clearly positively invariant.

5.15.1 Monotonicity

We consider the Jacobian of this system

$$\mathrm{Jac} = \begin{bmatrix} -\mu_u & \lambda & 0 & 0 \\ r - \frac{r\, A_w^2}{(A_w + A_u)^2} & -\lambda - \nu - d\, L_w - 2\, d\, L_u & -\frac{r\, A_u^2}{(A_w + A_u)^2} & -d\, L_u \\ 0 & 0 & -\mu_w & \lambda \\ 0 & -d\, L_w & r & -\lambda - \nu - 2d\, L_w - d\, L_u \end{bmatrix}$$

If we consider the permutation matrix

$$P = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix $P\,\mathrm{Jac}\,P$ is aMetzler matrix. This means that the system is monotone for the cone $K = -\mathbb{R}^+ \times -\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$.

5.15.2 Strong monotonicity

The system is irreducible. If we draw the graph of the Jacobian we obtain

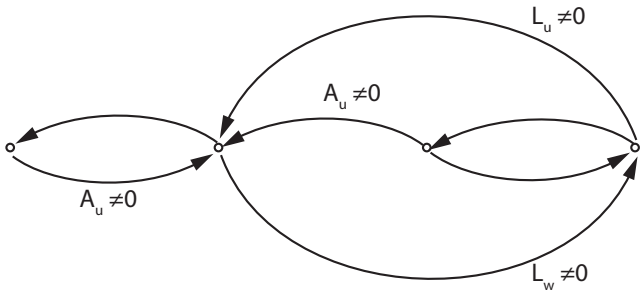


Figure 5.1: Graph of Jacobian

The system will be strongly monotone if $L_w \neq 0$ and $A_u \neq 0$.

5.15.3 Equilibria

We have three equilibria.

A DFE

$$L_u^* = \frac{(r\lambda - \mu_u(\lambda+\nu))}{d\,\mu_u} \qquad A_u^* = \frac{\lambda\,(r\lambda - \mu_u(\lambda+\nu))}{d\,\mu_u^2}$$

This equilibrium exists if the basic offspring number, for uninfected, satisfies

$$\mathcal{R}_{\text{offsp},u} = \frac{r\,\lambda}{\mu_u\,(\lambda+\nu)} > 1$$

A *Wolbachia* completely infected equilibrium (WCIE)

$$L_w^* = \frac{(r\lambda - \mu_w(\lambda+\nu))}{d\,\mu_w} \qquad A_w^* = \frac{\lambda\,(r\lambda - \mu_w(\lambda+\nu))}{d\,\mu_w^2}$$

This equilibrium exists if the basic offspring number, for infected, satisfies

$$\mathcal{R}_{\text{offsp},w} = \frac{r\,\lambda}{\mu_i\,(\lambda+\nu)} > 1$$

A coexistence equilibrium

$$\bar{L}_u = \frac{\mu_u^2\,[r\,\lambda - \mu_w\,(\lambda+\nu)]}{d\,\mu_w\,(\mu_u^2 - \mu_w\,\mu_u + \mu_w^2)} \qquad \bar{A}_u = \frac{\lambda\,[r\,\lambda - \mu_u\,(\lambda+\nu))}{d\,\mu_u^2}$$
$$\bar{L}_w = \frac{(\mu_w - \mu_u)\,[r\,\lambda - \mu_w\,(\lambda+\nu)]}{d\,(\mu_u^2 - \mu_w\,\mu_u + \mu_w^2)} \qquad \bar{A}_w = \frac{\lambda\,[r\lambda - \mu_w(\lambda+\nu)]}{d\,\mu_w^2}$$

This equilibrium exists if $\mathcal{R}_{\text{offsp},w} > 1$

Since *Wolbachia* is life shortening we assume that $\mu_w > \mu_u$. Hence if $\mathcal{R}_{\text{offsp},w} > 1$ we have three equilibria.

5.15.4 Basic reproduction ratio

We will now consider the basic reproduction ratio for the infection $\mathcal{R}_0$.
The Jacobian at the DFE is

$$J(A_u^*, L_u^*, 0, 0) = \begin{bmatrix} -\mu_u & \lambda & 0 & 0 \\ r & -\frac{2r\lambda}{\mu_u} + \lambda + \nu & -r & -\frac{r\lambda}{\mu_u} + \lambda + \nu \\ 0 & 0 & -\mu_i & \lambda \\ 0 & 0 & r & -\frac{r\lambda}{\mu_u} \end{bmatrix}$$

This is a block upper triangular matrix. The upper block diagonal is

$$J_1 = \begin{bmatrix} -\mu_u & \lambda \\ r & -\frac{2r\lambda}{\mu_u} + \lambda + \nu \end{bmatrix}$$

The trace of this matrix is $-\frac{r\lambda}{\mu_u} + \lambda + \nu - \mu_u - \frac{r\lambda}{\mu_u} < 0$, if $\mathcal{R}_{\text{offsp},u} > 1$. The determinant of J_1 is $r\lambda - \mu_u(\lambda+\nu) > 0$. This block is Hurwitz. Therefore hypothesis **H5** is satisfied.
We can now look at the Jacobian of the transmission in the infected compartment

$$F = \begin{bmatrix} 0 & 0 \\ r & 0 \end{bmatrix}$$

The Jacobian of the other terms is

$$V = \begin{bmatrix} -\mu_i & \lambda \\ 0 & -\frac{r\lambda}{\mu_u} \end{bmatrix}$$

Hence the next generation matrix is

$$K = \begin{bmatrix} 0 & 0 \\ \frac{r}{\mu_w} & \frac{\mu_u}{\mu_w} \end{bmatrix}$$

and

$$\mathcal{R}_0 = \frac{\mu_u}{\mu_w} < 1$$

Which proves the asymptotic stability of the DFE.

5.15.5 Stability of the CWIE

Consider the Jacobian at the CWIE

$$J(0,0,A_w^*,L_w^*) = \begin{bmatrix} -\mu_u & \lambda & 0 & 0 \\ 0 & -\frac{r\lambda}{\mu_w} & 0 & 0 \\ 0 & 0 & -\mu_w & \lambda \\ 0 & -\frac{r\lambda}{\mu_w} + \lambda + \nu & r & -\frac{2r\lambda}{\mu_w} + \lambda + \nu \end{bmatrix}$$

This a lower block triangular matrix. The lower diagonal block is

$$J_4 = \begin{bmatrix} -\mu_w & \lambda \\ r & -\frac{2r\lambda}{\mu_w} + \lambda + \nu \end{bmatrix}$$

which is clearly Hurwitz if $\mathcal{R}_{\text{offsp},w} > 1$.
The upper diagonal block is

$$J_1 = \begin{bmatrix} -\mu_u & \lambda \\ 0 & -\frac{r\lambda}{\mu_w} \end{bmatrix}$$

which is Hurwitz without any condition.
The Jacobian $J(CWIE)$ is then Hurwitz, which proves the asymptotic stability of the CWIE.

5.15.6 Global analysis

We have 3 equilbria, we will use Hirsch's Theorem (5.11.1). We must prove that the trajectories are forward bounded

Solutions are forward bounded

we denote $X_u = (A_u, L_u)$ and $X_w = (A_w, L_w)$. The vector field on $\mathbb{R}^4_+$ defined by (5.15) will be denoted by $f(A_u, L_u, A_w, L_w) = (f_u(A_u, L_u, A_w, L_w), f_w(A_u, L_u, A_w, L_w))$ on $\mathbb{R}2_+ \times \mathbb{R}^2_+$.
The equilibria will be denoted by

$$X_{\mathrm{DFE}} = (A^*_u, L^*_u, 0, 0) \quad X_{\mathrm{WCIE}} = (0, 0, A^*_w, L^*_w) \quad X_{\mathrm{coex}} = (\bar{A}_u, \bar{L}_u, \bar{A}_w, \bar{L}_w,)$$

and

$$X^*_u = (A^*_u, L^*_u) \quad X^*_w = (A^*_w, L^*_w)$$

Recall that the order is given by the cone $K = -\mathbb{R}^+ \times -\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$. We denote by $\leq$ the classical order given by the nonnegative orthant and by $\leq_K$ the order associated to K.
We have

$$X_{\mathrm{DFE}} \ll_K X_{\mathrm{coex}} \ll_K X_{\mathrm{WCIE}}$$

Since $f(X_{\mathrm{DFE}}) = 0$ and $f(X_{\mathrm{WCIE}}) = 0$, by proposition (5.8.2) the order interval (for K-order) $[X_{\mathrm{DFE}}, X_{\mathrm{WCIE}}]_K$ is positively invariant.
We can obtain more : if $\xi > 1$ we have $f_u(\xi\, A^*_u, \xi\, L^*_u, 0, 0) < \xi\, f_u(A^*_u,\, L^*_u, 0, 0) = 0$ and $f_w(0, 0, \xi\, A^*_w, \xi\, L^*_w) < \xi\, f_u(0, 0, A^*_w,\, L^*_w) = 0$. For the K-order

$$\xi\, X_{\mathrm{DFE}} = (\xi\, A^*_u, \xi\, L^*_u, 0, 0) \ll_K \xi\, X_{\mathrm{WCIE}} = (0, 0, \xi\, A^*_w, \xi\, L^*_w) \text{ with}$$
$$f(\xi\, X_{\mathrm{DFE}}) >_K 0 \qquad \text{and} \qquad f(\xi\, X_{\mathrm{WCIE}}) <_K 0$$

This proves by proposition (5.8.2) that $[\xi\, X_{\mathrm{DFE}}, \xi\, X_{\mathrm{WCIE}}]_K$ is positively invariant. For ξ large enough any element of the nonnegative orthant can be included in this order interval. Therefore any trajectory is bounded.

Stability analysis

We consider the order interval $[X_{\mathrm{DFE}}, X_{\mathrm{coex}}]_K$ and the 2-face defined by $A_w = L_w = 0$. The 2-face is positively invariant and we claim that X_{DFE} is GAS on this face. It is sufficient to note that this system on the 2-face is strongly monotone, strictly sublinear and that X^*_u is asymptotically stable. By Hirsch theorem (5.10.1) X^*_u is GAS. We know that X_{DFE} is asymptotically stable because $\mathcal{R}_0 > 1$. Hence all the hypothesis of Theorem (5.11.1) are satisfied. All trajectories in $[X_{\mathrm{DFE}}, X_{\mathrm{coex}}]_K$

converge to X_{DFE} and X_{coex} is unstable. This result is obtained without any computation.
Actually $[\xi\, X_{\mathrm{DFE}}, X_{\mathrm{coex}}]_K$ is in the basin of attraction of X_{coex}. Left as an exercise.

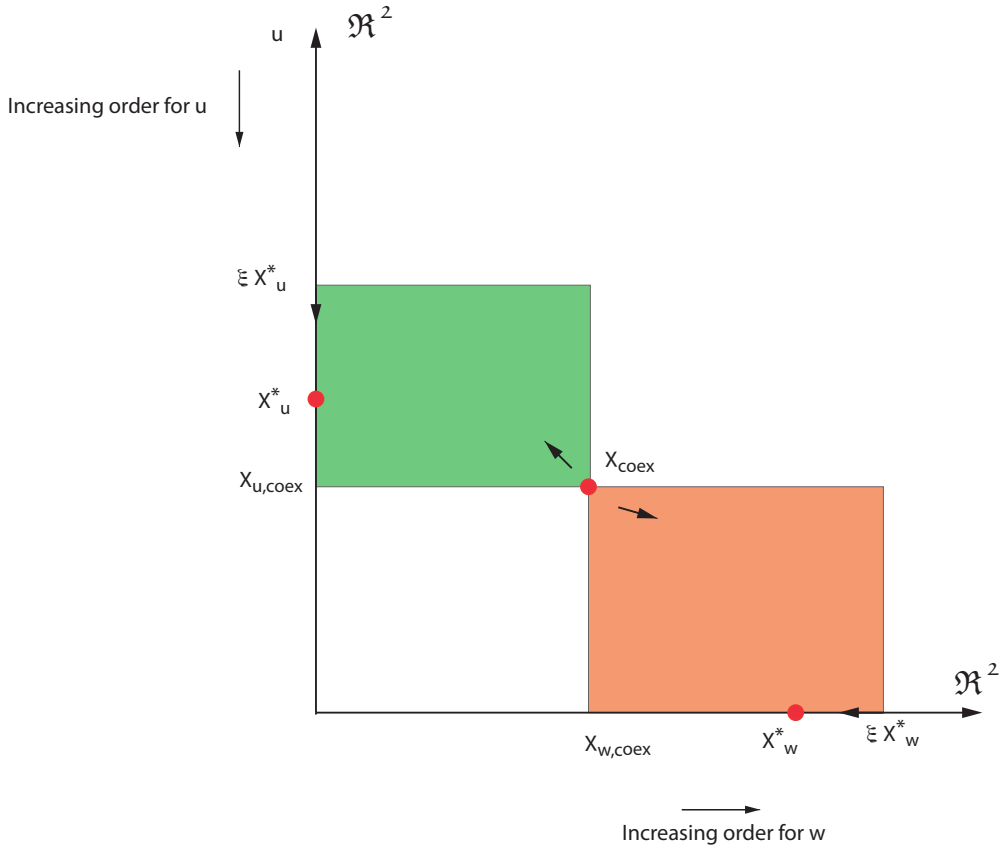


Figure 5.2: Order interval and equilibria

5.16 Brucellosis

We consider the following model for ovine brucellosis incorporating direct and indirect transmission from

B. Ainseba, C. Benosman and P. Magal

Journal of Biological Dynamics, vol 4, n¡ 1, 2010 pp 2-11

This section left as an exercise has for objective to give easily the stability analysis of the model. Compare with the original paper !

$$\begin{cases} \dot{S} &= b\,S - \left(m + \dfrac{(b-m)}{K}\,N\right)\,S + (1-p)\,b\,I - a_1\,S\,I - a_2\,S\,C \\ \dot{I} &= p\,b\,I - \left(m + \dfrac{(b-m)}{K}\,N\right)\,I + a_1\,S\,I + a_2\,S\,C \\ \dot{C} &= k_1\,I\,(1-C)\,-k_2,C \end{cases}$$

Brucellosis is due to a virus *Brucella*. It is transmitted from animals to humans either by ingestion of contaminated products, such as milk or vegetables cultivated on soil containing contaminated manure, or directly via the mucosa upon contact with the infected organisms. This model incorporate vertical transmission and the contamination of the environment.
AS usual $S(t)$ and $I(t)$ are the susceptible and infected at time t and $C(t)$ is the fraction of contaminated environment. p is the proportion of newborns that are infected. The population dynamic is described by a logistic equation with b the birth rate and m the death rate. It assumed that $b \geq m$.
Show that the compact

$$\mathcal{K} = \{(S,I,C) \in \mathbb{R}^3_+ \mid \ 0 \leq S+I \leq N \ \ 0 \leq C \leq 1$$

is positively invariant.
Hence use Vidyasagar's Theorem (3.6.1) to show that the stability of the following system, on the domain $[0,K] \times [0,1]$, is equivalent to the stability of the 3-dimensional system

$$\begin{cases} \dot{I} &= -b\,(1-p)\ I + a_1\,(K-I)\,I + a_2\,(K-I)\,C \\ \dot{C} &= k_1\,I\,(1-C)\,-k_2\,C \end{cases}$$

Show that the system is strongly monotone on $\mathcal{K}$
Show that the right side of the ODE is strongly sublinear on $\mathcal{K}$.
Show that

$$\mathcal{R}_0 = \frac{a_1\,K}{b\,(1-p)} + \frac{a_2\,k_1\,K}{k_2\,b\,(1-p)}$$

Use Hirsch's theorem (5.10.1 to show that
if $\mathcal{R}_0 < 1$ the DFE is GAS on $\mathcal{K}$
If $\mathcal{R}_0 > 1$ there exists an unique EE in $\mathcal{K}$ which is GAS.
If $\mathcal{R}_0 = 1$, consider the candidate Lyapunov function $V(I,C) = k_2\,I + a_2\,K\,C$. Show that the origin is stable.

Use again Hirsch's Theorem Hirsch's theorem (5.10.1 to exclude an endemic EE and to the stability of the DFE. You can also use LaSalle's principle.

5.17 Population dynamics of mosquito

The life cycle of a mosquito consists of two main stages: aquatic (egg, larva, pupa) and adult (with males and females). After emergence from pupa, a female mosquito needs to mate and get a blood meal before it starts laying eggs. Then every $4--5$ days it will take a blood meal and lay $100--150$ eggs at different places ($10--15$ per place). For the mathematical description, our model is inspired by the model considered in [29, 5].
However we will consider three aquatic stages, where the authors [29, 5] lump the three stages into a single aquatic stage. The rationale is to prepare for a subsequent model with infection by *Wolbachia*. Furthermore, we split the adult stage into three sub-compartments, males, immature female and mature female which leads to the following compartments:

- Eggs E;
- Larvae L;
- Pupae P;
- Males M;
- Young immature females Y; We consider a female to be in the Y compartment from its emergence from pupa until her gonotrophic cycle has began, that is the time of mating and taking the first blood meal, which takes typically $3--4$ days.
- Mature females F, i.e., fertilized female. A female needs to mate successfully only once and rarely remate.

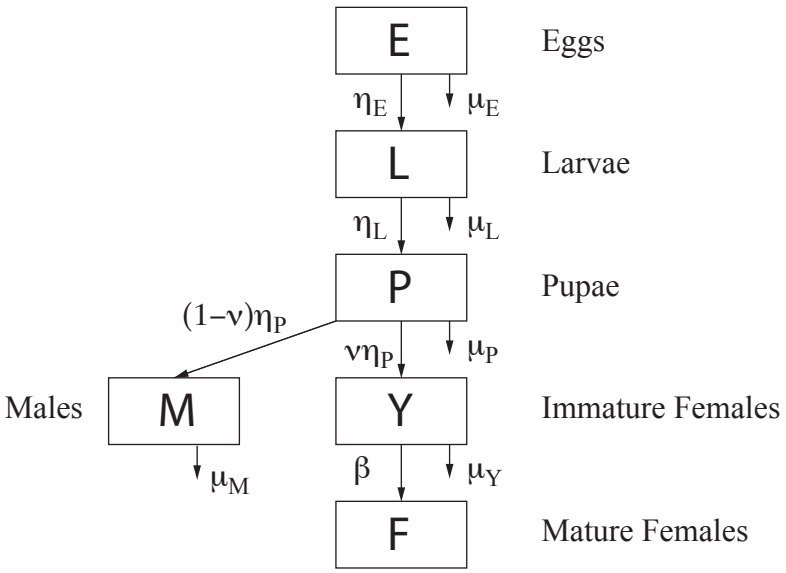
The parameters μ_E, μ_P, μ_Y, μ_F and μ_M are respectively the death rate of eggs, larvae, pupae, immature female, mature females and males. The parameters η_E,η_L, η_P , β are the respective rate of transfer to the next compartment. The parameter ν is the sex ratio. In this model, we use a density dependent death rate for the larvae stage since mosquitoes larvae (anopheles and aedes) are density sensitive, which imply an additional density mortality rate $\mu_2 L$. The equation for L can be considered as a logistic equation. Such an hypothesis is appropriate since mosquitoes only have access to a finite number of potential breeding sites, and density-dependent larval survival has been demonstrated at such sites. The

parameter ϕ is the average amount of eggs laid per fertilized female per unit of time.

Mating is a complex process that is not fully understood. However, as discussed in [5] and references therein, the male mosquito can mate practically through all its life. A female mosquito needs one successful mating to breed lifelong. It is admitted that mosquitoes locate themselves in space and time to ensure they are available to mate. Therefore, it is reasonable to assume that in any case the immature female will mate and afterwards move to compartment F , or die. Thus a parameter like $\frac{1}{\beta+\mu_Y}$ can represents the mean time given by length of the first gonotrophic cycle of a female, i.e., the interval from immediately after the mating to the first blood meal.

We assume that all the parameters are constant. In reality, the mosquito population varies seasonally. Nevertheless, such a model should be a good approximation for a definite season.

5.17.1 The model



$$\begin{cases} \dot{E} &= \phi\, F - (\mu_E + \eta_E)\, E \\ \dot{L} &= \eta_E\, E - (\mu_L + \eta_L + \mu_2\, L)\, L \\ \dot{P} &= \eta_L\, L - (\mu_P + \eta_P)\, P \\ \dot{Y} &= \nu\, \eta_P\, P - (\beta + \mu_Y)\, Y \\ \dot{F} &= \beta\, Y - \mu_F\, F \\ \dot{M} &= (1-\nu)\, \eta_P\, P - \mu_M\, M. \end{cases} \tag{5.16}$$

If we denote by X a vector of the state space of this systems.

$$X^T = (E, L, P, Y, F, M),$$

then the systems can be written

$$\dot{X} = A(X)\, X,$$

For (5.16) the matrix is given by

$$A(X) = \begin{bmatrix} -(\mu_E + \eta_E) & 0 & 0 & 0 & \phi & 0 \\ \eta_E & -(\mu_L + \eta_L + \mu_2\, L) & 0 & 0 & 0 & 0 \\ 0 & \eta_L & -(\mu_P + \eta_P) & 0 & 0 & 0 \\ 0 & 0 & \nu\, \eta_P & -(\beta + \mu_Y) & 0 & 0 \\ 0 & 0 & 0 & \beta & -\mu_F & 0 \\ 0 & 0 & (1-\nu)\, \eta_P & 0 & 0 & -\mu_M \end{bmatrix}.$$

The matrix $A(X)$ is a Metzler matrix, this implies that the nonnegative orthant is positively invariant for (5.16).
Actually $A(X)$ depends only of L, so we will denote the matrix $A(X)$ simply by $A(L)$.

5.17.2 Analysis of the model

We can define a basic offspring number, using the techniques of $\mathcal{R}_0$, where the transmission term is given by $\phi\, F$) :
Using (5.16) shows that

$$\mathcal{R}_{0,\text{offsp}} = \frac{\phi}{\mu_F}\,\frac{\eta_E}{\mu_E+\eta_E}\,\frac{\eta_L}{\mu_L+\eta_L}\,\frac{\nu\,\eta_P}{\mu_P+\eta_P}\;\;\frac{\beta}{\beta+\mu_Y}.$$

When $\mathcal{R}_{0,\text{offsp}} \leq 1$ the only equilibrium is the origin. When $\mathcal{R}_{0,\text{offsp}} > 1$ a second positive equilibrium exists $X^* = (E^*, L^*, P^*, Y^*, F^*, M^*)^T$.
We can express all the components as positive linear expressions of P^*

$$L^* = \frac{\mu_p+\eta_p}{\eta_L}\,P^*, \qquad Y^* = \frac{\nu\,\eta_P}{\beta+\mu_Y}\,P^*, \tag{5.17}$$

$$F^* = \frac{\beta}{\beta+\mu_Y}\,\frac{\nu\,\eta_P}{\mu_F}\,P^*, \qquad M^* = \frac{(1-\nu)\,\eta_P}{\mu_M}\,P^* \tag{5.18}$$

$$E^* = \frac{\phi}{\mu_E+\eta_E}\,\frac{\beta}{\beta+\mu_Y}\,\frac{\nu\,\eta_P}{\mu_F}\,P^*. \tag{5.19}$$

Finally, replacing in the equation $\dot{L} = 0$, we get

$$\begin{aligned} P^* &= \frac{\eta_L\,(\mu_L+\eta_L)}{\mu_2\,(\mu_P+\eta_P)}\,\left(\frac{\eta_E\,\nu\,\phi\,\eta_L\,\eta_P}{\mu_F\,(\mu_E+\eta_E)\,(\mu_L+\eta_L)\,(\mu_P+\eta_P)}-1\right)\\ &= \frac{\eta_L\,(\mu_L+\eta_L)}{\mu_2\,(\mu_P+\eta_P)}\,(\mathcal{R}_{0,\text{offsp}}-1) > 0. \end{aligned} \tag{5.20}$$

The system is monotone

$$\text{Jac} = \begin{bmatrix} -\eta_E-\mu_E & 0 & 0 & 0 & \phi & 0\\ \eta_E & -\eta_L-\mu_2-\mu_L & 0 & 0 & 0 & 0\\ 0 & \eta_L & -\eta_P-\mu_P & 0 & 0 & 0\\ 0 & 0 & \nu\,\eta_P & -\beta-\mu_Y & 0 & 0\\ 0 & 0 & 0 & \beta & -\mu_F & 0\\ 0 & 0 & (1-\nu)\,\eta_P & 0 & 0 & -\mu_M \end{bmatrix}$$

Then by proposition (5.8.2) the order interval $[0, x^*]$ is positively invariant.
The system is strictly sublinear. Denote by $T(X) = A(X)\,X$, then for $0 < \lambda < 1$ and $X \gg 0$

$$\lambda\,T(X) = \lambda\,A(L)\,L < \lambda\,A(\lambda\,L)\,L = T(\lambda\,X)$$

Let now $\lambda > 1$, then by sublinearity $\frac{1}{\lambda} T(\lambda X^*) < T(X^*) = 0$. Therefore the order interval $[0, \lambda X^*]$ is positively invariant by proposition (5.8.2). We conclude that all the trajectories are forward bounded. Then we can us Vidyasagar's Theorem to reduce the system, suppressing the equation for the males.
The Jacobian for the reduced system is now

$$\begin{bmatrix} -\eta_E - \mu_E & 0 & 0 & 0 & \phi \\ \eta_E & -\eta_L - \mu_2 - \mu_L & 0 & 0 & 0 \\ 0 & \eta_L & -\eta_P - \mu_P & 0 & 0 \\ 0 & 0 & \nu\,\eta_P & -\beta - \mu_Y & 0 \\ 0 & 0 & 0 & \beta & -\mu_F \end{bmatrix}$$

This matrix is clearly irreducible and the system is strongly monotone and strictly sublinear. We conclude with Hirsch's Theorem

Theorem 5.17.1
If the basic reproduction ratio $\mathcal{R}_{0,offsp} > 1$ then there is an unique positive equilibrium X^* which is GAS on $\mathbb{R}^6_+ \setminus \{0\}$.

5.18 A schistosomiasis model

We consider the following model from
Allen, E J and Victory, H D Jr, *Modelling and simulation of a schistosomiasis infection with biological control.* Acta Tropica, 2 vol 87, pp 251–267 (2003).

We consider the following system with variables :

- I_h infected human by schistosomes
- E_s the latent snails
- I_s infectious snails releasing cercariae
- I_m infected mammals
- N_h the human population
- N_s the snail population for transmission for schistosomiasis
- N_{rs} a resistant snail population, competitor for the first population
- N_m a mammals population.

Two subsystem. A demographic model

$$\begin{cases} \dot{N}_h &= \Lambda - \mu_h\, N_n \\ \dot{N}_s &= r_s\, N_s\, \left(1 - \dfrac{N_s}{K_s} - \lambda_{rs}\, N_{rs}\right) \\ \dot{N}_{rs} &= r_{rs}\, N_{rs}\, \left(1 - \dfrac{N_{rs}}{K_{rs}} - \lambda_s\, N_s\right) \\ \dot{N}_m &= r_m\, N_m\, \left(1 - \dfrac{N_m}{K_m}\right) \end{cases}$$

and a transmission model

$$\begin{cases} \dot{I}_h &= \beta_{hs}\, (N_h - I_h)\, I_s - \gamma_h\, I_h \\ \dot{E}_s &= (\beta_{sh}\, I_h + \beta_{sm}\, I_m)\, (N_s - E_s - I_s) - (\mu_s + \alpha_s + \dfrac{N_s}{K_s} + \lambda_{rs}\, N_{rs})\, E_s \\ \dot{I}_s &= \alpha_s\, E_s - (\mu_s + \dfrac{N_s}{K_s} + \lambda_{rs}\, N_{rs})\, I_s \\ \dot{I}_m &= \beta_{ms}\, (N_m - I_m)\, I_s - (\mu_m + \dfrac{r_m}{K_m}\, N_m)\, I_m \end{cases}$$

It is immediate that the nonnegative orthant is positively invariant. By considering the population dynamics equations, the compact

$$K = \{X \in \mathbb{R}^8_+ | N_h \leq \frac{\Lambda}{\mu_h}, N_s \leq K_s, N_{rs} \leq K_{rs}, N_m \leq K_m\},$$

is positively invariant and absorbing. Absorbing means that any trajectory tends to K when $t \to +\infty$. Therefore any trajectory is forward bounded.
Considering the population dynamics it is clear that $N_h^* = \dfrac{\Lambda}{\mu_h}$ and $N_m^* = K_m$ are GAS equilibria. For the equations giving (N_s, N_m) we have a classical Lotka-Volterra competition model. This is analyzed for example in [48, 98].
We assume that $\lambda_s\, K_s > 1$ and $\lambda_{rs}\, K_{rs} < 1$. In this case the positive coexistence equilibrium exists (N_s^*, N_{rs}^*) and is GAS on the interior of the nonnegative orthant. We can apply Vidyasagar's Theorem to obtain a reduced system.

$$\begin{cases} \dot{I}_h &= \beta_{hs}\,(N_h^* - I_h)\,I_s - \gamma_h\,I_h \\ \dot{E}_s &= (\beta_{sh}\,I_h + \beta_{sm}\,I_m)\,(N_s^* - E_s - I_s) - (\gamma_s + \alpha_s)\,E_s \\ \dot{I}_s &= \alpha_s\,E_s - \gamma_s\,I_s \\ \dot{I}_m &= \beta_{ms}\,(N_m^* - I_m)\,I_s - \gamma_m\,I_m \end{cases}$$

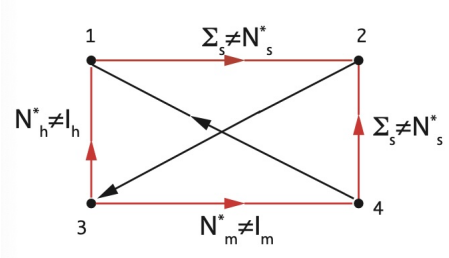
With a variable change, with new variables $(I_h, \Sigma_s = N_s^* - S_s, I_s, I_m)$, the system is equivalent to

$$\begin{cases} \dot{I}_h &= \beta_{hs}\,(N_h^* - I_h)\,I_s - \gamma_h\,I_h \\ \dot{\Sigma}_s &= -\gamma_s\,N_s^* + (\beta_{sh}\,I_h + \beta_{sm}\,I_m + \gamma_s)\,(N_s^* - \Sigma_s) \\ \dot{I}_s &= \alpha_s\,(\Sigma_s - I_s) - \gamma_s\,I_s \\ \dot{I}_m &= \beta_{ms}\,(N_m^* - I_m)\,I_s - \gamma_m\,I_m \end{cases} \tag{5.21}$$

The jacobian is

$$\begin{bmatrix} -\beta_{sh}\,I_s - \gamma_h & 0 & \beta_{hs}\,(N_h^* - I_h) & 0 \\ \beta_{sh}\,(N_S^* - \Sigma_s) & -(\beta_{sh}\,I_h + \beta_{sm}\,I_m + \gamma_s) & 0 & \beta_{sm}\,(N_S^* - \Sigma_s) \\ 0 & -\alpha_s & -(\alpha_s + \gamma_s) & 0 \\ 0 & 0 & \beta_{ms}\,(N_m^* - I_m) & -\beta_{ms}\,I_s - \gamma_m \end{bmatrix}$$

The Jacobian is a Metzler matrix. If we look for the graph of this matrix



The system is strongly monotone excepted in some hyperplane defined by $\Sigma_s = N_s^*$,i.e., all the snail are infected and/or infectious and by $I_m = N_m^*$, i.e., all the mammal are infected. These hyperplanes are not positively invariant and any trajectory, starting in one of these surfaces, leaves these hyperplanes.

Left as an exercise, we compute $\mathcal{R}_0$ for the DFE $(0, 0, 0, 0)$

$$\mathcal{R}_0^2 = \frac{\alpha_s\,\gamma_m\,\beta_{\rm hs}\,\beta_{\rm sh}\,N_h^*\,N_s^* + \alpha_s\,\gamma_h\,\beta_{\rm sm}\,\beta_{\rm ms}\,N_m^*\,N_s^*}{\gamma_h\gamma_m\alpha_s\gamma_s + \gamma_h\gamma_m\gamma_s^2}.$$

It is not difficult to check that the system is strictly sublinear.
We have, denoting by F the vector field defined by the system (5.21)

$$F(N_h^*, N_s^*, N_s^*, N_m^*) = (-\gamma_n\,N_h^*, -\gamma_s\,N_s^*, -\gamma_s\,N_s^*, -\gamma_m\,N_m^*)^T \ll 0$$

Hence by Hirsch's Theorem if $\mathcal{R}_0 > 1$ there exists a unique positive equilibrium which is GAS on the nonnegative orthant minus the origin. We can also use the proposition using sublinearity.

5.19 A metapopulation model with a disease

We begin to study the demographic model.
The demographic model is the combination of a demographic process on each patch with a migration process. We use, for the migration process, the framework developed by Arino and van den Driessche[6, 7]. We consider population dynamics (deaths or births) within the patches. The population birth rate is analogous to the function used in [8, 99]. These birth rate encompasses such laws as Ricker, Beverton-Holt and constant recruitment. Our model can also represent single-species dynamical system which is composed of several patches connected by discrete diffusion like in [70, 89].

5.19.1 The demographic model

Recruitment

Following [20], we consider a family of functions $\mathbb{B}(x)$. These functions represent the per capita birth rate.
The dynamics of demography are governed by $\dot{x} = \mathbb{B}(x) - \mu\,x$, where $\mu\,x$ is the per capita death rate. We assume that this family satisfies the following hypotheses. Each function $\mathbb{B}$ is differentiable on $\mathbb{R}$ and there exists a continuously differentiable function $\mathcal{B}(x)$ on $(0, +\infty)$ such that for any $x > 0$:

- **[H1]** $\mathbb{B}(x) = \mathcal{B}(x)\,x$

- **[H2]** $\mathcal{B}(x) > 0$
- **[H3]** $\mathcal{B}'(x) < 0$
- **[H4]** $\mu > \mathcal{B}(+\infty)$

In [20] a supplementary hypothesis is assumed : **[H5]** $\mathcal{B}(0+) > \mu$. This hypothesis ensures that there is no natural extinction of the population. We will relax this hypothesis: we will consider the family of functions composed of functions of type $\mathbb{B}$ satisfying the hypothesis **[H1]** to **[H4]**, to which we also add the zero function. These assumptions are used to take into account the possibility of the absence of reproduction in some places and eventually extinction when there is no mobility. Then the family of functions $\mathbb{B}$ is composed of functions such that $\mathbb{B}(0) = 0$, satisfying hypotheses **[H1]** to **[H4]**, together with the constant nonnegative functions. When $\mathbb{B}(0) = 0$, since $\mathbb{B}$ is $\mathcal{C}^1$, the function $\mathcal{B}$ satisfying hypothesis **[H1]** is continuous on $\mathbb{R}_+$, thus $\mathcal{B}(0)$ makes sense.

The dynamics, given by $\dot{x} = \mathbb{B}(x) - \mu\, x$, satisfy:

- the half-line $\mathbb{R}_+$ is positively invariant,
- if $\mathbb{B}$ is not the zero function and **[H5]** is satisfied, then there exists a unique positive demographic equilibrium $\bar{x}$, which is globally asymptotically stable on $\mathbb{R}^+ \setminus \{0\}$,
- if **[H5]** is not satisfied, the origin is globally asymptotically stable.

The considered family includes for instance:

- the constant recruitment functions: $\mathbb{B}(x) = \Lambda$;
- the Ricker type functions [92, 20] $\mathcal{B}(x) = \alpha\, e^{-\beta\, x}$, with $\alpha > \mu > 0,\ \beta > 0$;
- the Beverton-Holt functions : [92, 20] $\mathcal{B}(x) = \frac{\alpha}{1+\beta\, x^m}$, with $\alpha > \mu > 0,\ \beta > 0,\ m > 0$;
- Deriso-Schnute functions: $\mathcal{B}(x) = \alpha\,(1 - \gamma\beta x)^{1/\gamma}$, with $\alpha > \mu > 0,\ \beta > 0,\ \gamma > 0$.

The migration process

Arino-van den Driessche migration model

We consider n patches. This model does not keep track of where an individual usually resides, but just considers where he is at time t. The transfer rate from patch i to patch j, for $i \neq j$, is denoted by $m_{ji} \geq 0$. The total host population in patch i is denoted by N_i. The per capita birth rate is given by $\mathcal{B}_i(N_i)\, N_i$, the per capita death rate by $\mu_i\, N_i$. Hence, for $i = 1, \cdots, n$, the dynamics is given by

$$\dot{N}_i = \mathcal{B}_i(N_i)\, N_i - \mu_i\, N_i + \sum_{j=1\, j\neq i}^{n} m_{ij}\, N_j - N_i \sum_{j=1\, j\neq i}^{n} m_{ji}.$$

This system can be written in a condensed form

$$\dot{N} = \mathrm{diag}(\mathbb{B}(N)) - \mathrm{diag}(\mu)\, N + M\, N = \mathrm{diag}(\mathcal{B}(N))\ N - \mathrm{diag}(\mu)\, N + M\, N. \quad (5.22)$$

Where N is the column vector $(N_1, \cdots, N_n)^T$, the superscript T denoting transpose, the matrix $\mathrm{diag}(\mathcal{B}(N)$ denotes the diagonal matrix whose diagonal elements are given by the vector

$$\mathcal{B}(N) = (\mathcal{B}_1(N_1), \cdots, \mathcal{B}_n(N_n))^T,$$

and $\mathrm{diag}(\mu)$ denotes the diagonal matrix whose diagonal elements are given by the

$$\mu = (\mu_1, \cdots, \mu_n)^T.$$

the matrix M is defined by $M(i,j) = m_{i,j}$ for $i \neq j$ and

$$M(i,i) = -\sum_{j=1\, j\neq i}^{n} m_{ji}\ ,$$

We also define, for further reference the matrix $\mathrm{diag}(\mathbb{B}(N))$.
As we have already seen the matrix M is a Metzler matrix.
Let $\mathbb{1}$ be the vector of $\mathbb{R}^n : \mathbb{1} = (1, \cdots, 1)^T$. Each column sum of the matrix M is zero. This can be written

$$\mathbb{1}^T\, M = 0.$$

Takeuchi diffusion model

According to [70] we have the model

$$\dot{N}_i = \mathcal{B}_i(N_i)\, N_i - \mu_i\, N_i + \sum_{j=1\, j\neq i}^{n} m_{ij}\, (N_j - N_i).$$

m_{ij} here is a nonnegative diffusion coefficient for the species from j-th patch to i-th patch. It is supposed in this model that the net exchange from the j-th patch to the i-th patch is proportional to the difference $(N_j - N_i)$ of population densities in each patch.
In the same vein we define by $M(i,j) = m_{i,j}$ for $i \neq j$ and

$$M(i,i) = -\sum_{j=1\, j\neq i}^{n} m_{ij} \ ,$$

again M is Mezler matrix, but with the difference $M.\mathbb{1} = 0$.
The model is also represented by system (5.22).

Stability Analysis

Theorem 5.19.1
On the nonnegative orthant, we consider the system (5.22). We further assume that the matrix M is irreducible.
If either

- $\mathbb{B}(0) = 0$ *and there is at least one function $\mathcal{B}_i(x)\,x$ which is not zero and moreover if the stability modulus satisfies $s\Big(diag(\mathcal{B}(0)) - diag(\mu) + M\Big) > 0$;*
- *or* $\mathbb{B}(0) > 0$.

Then there exists an equilibrium $\bar{N} \gg 0$ which is globally asymptotically stable on the nonnegative orthant, except the origin.
Else $\mathbb{B}(0) = 0$ and $s\left(diag(\mathcal{B}(0)) - diag(\mu) + M\right) \leq 0$. Then the origin is globally asymptotically stable, i.e., the population goes to extinction in all patches.

Proof
We use Hirsch's Theorem (5.10.1).
First of all, the nonnegative orthant is positively invariant thanks to the assumptions on the matrix M and the functions $\mathcal{B}_i$.

We will show that there exists a positively invariant absorbing compact set for system (5.22):
Let I be the set of index i for which the function $\mathcal{B}_i(x)\,x$ is different from the null function, and let J be the set of the other indexes. One of these set can be empty.
Let H be the map defined by $H = \sum_{i=1}^{n} N_i$. We have

$$\dot{H} = \sum_{i\in I} \left(\mathcal{B}_i(N_i)\,N_i - \mu_i\,N_i\right) \ + \ \sum_{j\in J} \left(-\mu_j\,N_j\right) \ .$$

For each index $i \in I$, there exists $\tilde{N}_i > 0$ such that $B_i(\tilde{N}_i) \leq \mu_i$. Let us define $N^* = \max_i(\tilde{N}_i)$. Since $\mathcal{B}_i$ is decreasing, we have $\mathcal{B}_i(N^*) \leq \mu_i$. Thus, $\dot{H} \leq 0$ for $N \geq N^* \mathbb{1}$ and $\dot{H} < 0$ for $N > N^* \mathbb{1}$. This shows that the set $[0, N^* \mathbb{1}]$ is a positively invariant absorbing compact set. As a consequence all the trajectories of system (5.22) are forward bounded.

We claim that system (5.22) is cooperative and strongly monotone.
We set $F(N) = \text{diag}(\mathcal{B}(N))\ N - \text{diag}(\mu)\, N + M\, N$, i.e., the vector field associated to system (5.22). The derivative is given by

$$DF(N) = \text{diag}(\mathcal{B}(N)) + \text{diag}(\mathcal{B}'(N))\,\text{diag}(N) - \text{diag}(\mu) + M$$

The matrix M being Metzler it follows immediately that $DF(N)$ is Metzler, which proves that system (5.6) is monotone. Moreover since M is irreducible, system (5.6) is strongly monotone.

Using hypothesis **[H3]** and the fact that there is at least one index i such that $\mathcal{B}_i(N) > 0$, we have for $N \gg 0$,

$$F(N) - DF(N)\, N = -\text{diag}(\mathcal{B}'(N))\,\text{diag}(N)\, N > 0.$$

This proves the strict sublinearity of F by proposition (5.9.2).

If $\mathbb{B}(0) = 0$, then the origin is unstable since we have assumed in this case that $s\left(\text{diag}(\mathcal{B}(0)) - \text{diag}(\mu) + M\right) > 0$. Otherwise $\mathbb{B}(0) > 0$ and in this case the origin is not an equilibrium. Therefore, when the hypothesis (i) (respectively (ii)) of Theorem 5.19.1 is satisfied we have checked that system (5.6) satisfies all the conditions of Theorem (5.10.1), and hence, there exists an equilibrium $\bar{N} \gg 0$, which is globally asymptotically stable on $\mathbb{R}^n_+ \setminus \{0\}$.

To ends the proof of Theorem 5.19.1 it remains to consider the last case:

$$\mathbb{B}(0) = 0 \text{ and } s\Big(\text{diag}(\mathcal{B}(0)) - \text{diag}(\mu) + M\Big) \leq 0.$$

Since the matrix $\text{diag}(\mathcal{B}(0)) - \text{diag}(\mu) + M$ is an irreducible Metzler matrix, there exists $\mathbf{v} \gg 0$ such that

$$\mathbf{v}^T\ (\text{diag}(\mathcal{B}(0)) - \text{diag}(\mu) + M) = s\Big(\text{diag}(\mathcal{B}(0)) - \text{diag}(\mu) + M\Big)\,\mathbf{v}^T. \qquad (5.23)$$

We define the Lyapunov function on $\mathbb{R}^n_+$ by $V(N) = \mathbf{v}^T N$. The derivative $\dot{V}$ along the trajectories of system (5.6) is

$$\dot{V} = \mathbf{v}^T \Big(\text{diag}(\mathcal{B}(N)) - \text{diag}(\mu) + M\Big)\, N.$$

Since, by assumptions on the birth rate functions, $\mathcal{B}$ is decreasing and thanks to (5.23), we have

$$\dot{V} = \mathbf{v}^T \Big(\mathrm{diag}(\mathcal{B}(N)) - \mathrm{diag}(\mathcal{B}(0)) \Big)\, N + s \Big(\mathrm{diag}(\mathcal{B}(0)) - \mathrm{diag}(\mu) + M \Big)\, \mathbf{v}^T\, N \leq 0.$$

We consider the largest invariant set contained in $\dot{V} = 0$. This set is contained in the set $\{N | N_i = 0 \ i \in I\}$. Recall that I is the set of index i for which the function $\mathcal{B}_i(x)$ is different from the null function. If I is empty then F is simply a linear stable vector field. Otherwise by irreducibility of M, the largest invariant set contained in $\dot{V} = 0$ is reduced to the origin. This proves, by Lasalle's invariance principle the global asymptotic stability of the origin.

5.19.2 SIS disease

In this section we will study the stability of a metapopulation SIS model with a generalized law of contact.

The SIS model

Following [91] we consider the infection law given by $C(N)\, S \frac{I}{N}$, where N is the size of population, S and I respectively the number of susceptible and infectious individuals. $N = S + I$.

Diekman and Kretzschmar [27] suggest $C(N) = \dfrac{\zeta\, N}{1 + \alpha\, N}$, Anderson [3] propose $C(N) = \zeta\, N^{\alpha}$, Hesterbeek and Metz [38] use $\dfrac{\kappa_1\, N}{1 + \kappa_2\, N + \sqrt{1 + 2\,\kappa_2\, N}}$.

We denote by I_i, N_i respectively the number of infectious individuals and the total population in patch i. Taking into account the fact that $N_i = S_i + I_i$, the epidemic equation in patch i is given by

$$\dot{I}_i = C_i(N_i)\, (N_i - I_i)\, \frac{I_i}{N_i} - \mu_i\, I_i + \sum_{j=1\, j \neq i}^{n} m_{ij}\, I_j - I_i \sum_{j=1\, j \neq i}^{n} m_{ji}\ .$$

Using the same convention as in Section 5.19.1 we can write

$$\begin{cases} \dot{N} &= \mathrm{diag}(\mathcal{B}(N)\,)\, N - \mathrm{diag}(\mu)\, N + M\, N, \\ \dot{I} &= \mathrm{diag}(C(N))\, \mathrm{diag}(N)^{-1}\, \mathrm{diag}(N - I)\, I - \mathrm{diag}(\mu)\, I + M\, I. \end{cases} \tag{5.24}$$

Where I is the column vector $(I_1, \cdots, I_n)^T$ and $\mathrm{diag}(C(N))$ denotes the diagonal matrix whose diagonal elements are given by the vector

$$C(N) = (C_1(N_1), \cdots, C_n(N_n))^T, .$$

This system evolves on the closed positively invariant subset of $\mathbb{R}^{2n}$ given by $\mathbb{R}^n_+ \times \{I \mid 0 \leq I \leq N\}$.

Reduction of the model

System (5.24) is a triangular system, hence we can apply Vidyasagar's Theorem (3.6.1).

We have seen that when $s\left(\mathrm{diag}(\mathcal{B}(0)) - \mathrm{diag}(\mu) + M\right) > 0$, all the solutions of (5.6) are forward bounded and system (5.6) has a positive equilibrium, say $\bar{N} \gg 0$, which is globally asymptotically stable on $\mathbb{R}^n_+ \setminus \{0\}$. The solutions of (5.24) are forward bounded since they satisfy $I(t) \leq N(t)$ for all $t \geq 0$. Therefore, thanks to Vidyasagar's Theorem (3.6.1), the stability analysis of system (5.24) is reduced to the stability analysis of the following system

$$\dot{I} = \mathrm{diag}(C(\bar{N}))\,\mathrm{diag}(\bar{N})^{-1}\,\mathrm{diag}(\bar{N} - I)\,I - \mathrm{diag}(\mu)\,I + M\,I. \tag{5.25}$$

Computation of $\mathcal{R}_0$

In this section we will give an analytic expression for the basic reproduction ratio $\mathcal{R}_0$.

Proposition 5.19.1
The state $(\bar{N}, 0)$ is a disease-free equilibrium (DFE) of (5.24) and

$$\mathcal{R}_0 = \rho\left(-(M - diag(\mu))^{-1}\, diag(C(\bar{N}))\right).$$

Proof.
The Jacobian computed at the DFE, is $J((\bar{N}, 0)) = F + V$ where $F = \mathrm{diag}(C(\bar{N}))$ comes from infection and $V = M - \mathrm{diag}(\mu)$ is the part coming from other transfer. Since M is an irreducible Metzler matrix, with $s(M) = 0$, $M - \mathrm{diag}(\mu)$ is a nonsingular irreducible Metzler matrix. Hence $-(M - \mathrm{diag}(\mu))^{-1} \gg 0$ and thus

$$\mathcal{R}_0 = \rho(-F\,V^{-1}) = \rho\left(-(M - \mathrm{diag}(\mu))^{-1}\,\mathrm{diag}(C(\bar{N}))\right).$$

■

5.19.3 Existence and stability of equilibria

We can now state the stability result.

Theorem 5.19.2 *:*
We consider the system

$$\begin{cases} \dot{N} &= diag(\mathcal{B}(N)\,)\,N - diag(\mu)\,N + M\,N, \\ \dot{I} &= diag(C(N))\,diag(N)^{-1}\,diag(N-I)\,I - diag(\mu)\,I + M\,I \end{cases}$$

defined on $\Omega = \mathbb{R}^n_+ \times \{I \mid 0 \leq I \leq N\}$*, with* $s\,(diag(\mathcal{B}(0)) - diag(\mu) + M) > 0$ *and* M *irreducible.*
If $\mathcal{R}_0 \leq 1$*, then the disease free equilibrium is globally asymptotically stable (GAS).*

If $\mathcal{R}_0 > 1$*, then there exists a unique endemic equilibrium* $(\bar{N}, \bar{I}) \gg 0$*, which is GAS on* $\Omega \setminus \{(\bar{N}, 0)\}$*.*

Proof
We consider the vector field X of the reduced system (5.25), on the positively invariant set $\{0 \leq I \leq \bar{N}\}$.

$$X(I) = \mathrm{diag}(C(\bar{N}))\,\mathrm{diag}(\bar{N})^{-1}\,\mathrm{diag}(\bar{N} - I)\,I - \mathrm{diag}(\mu)\,I + M\,I.$$

Since

$$DX(I) = \mathrm{diag}(C(\bar{N}))\,\mathrm{diag}(\bar{N})^{-1}\,\mathrm{diag}(\bar{N} - 2I) - \mathrm{diag}(\mu) + M,$$

and since for $I \gg 0$ we have $X(I) - DX(I)\,I = \mathrm{diag}(I)\,I \gg 0$, the vector field is cooperative, irreducible and strongly sublinear. Hypothesis $\mathcal{R}_0 < 1$ implies that the origin is asymptotically stable. Hence by Theorem 5.10.1 the origin is globally asymptotically stable. Hypothesis $\mathcal{R}_0 > 1$ implies that the origin is unstable. This proves, using again Theorem 5.10.1, that there exists a unique equilibrium $\bar{I} \gg 0$ which is globally asymptotically stable.

It remains to consider the case $\mathcal{R}_0 = 1$. This implies $s(DX(0)) = 0$ and since $DX(0)$ is an irreducible Metzler matrix there exists a positive vector $\mathbf{v}$ such that

$$\mathbf{v}^T\,\left[\mathrm{diag}(C(\bar{N})) - \mathrm{diag}(\mu) + M\right] = 0$$

We use on $\{0 \leq I \leq \bar{N}\}$ the Lyapunov function $V(I) = \mathbf{v}^T\,I$. The derivative $\dot{V}$ along the trajectories is

$$\begin{aligned}\dot{V} &= \mathbf{v}^T \left[\operatorname{diag}(C(\bar{N})) - \operatorname{diag}(\mu) + M - \operatorname{diag}(C(\bar{N}))\operatorname{diag}(\bar{N})^{-1}\operatorname{diag}(I)\right] I \\ &= -\mathbf{v}^T \operatorname{diag}(C(\bar{N}))\operatorname{diag}(\bar{N})^{-1}\operatorname{diag}(I)\, I\end{aligned}$$

$\dot{V}$ is definite negative. This proves the global asymptotic stability of the origin. ■

5.20 Notes

To quote H. Thieme

> *The theory of (quasi-)positive matrices (i.e., Metzler) and of the associated dynamical system, as rarely as it is taught in standard linear algebra or ordinary differential equations courses, is a immensely powerful tool in population models with some kind of structures ...* H. Thieme pp 418 [92]

The Perron-Frobenius theorem appears everywhere in applied mathematics: in solutions using iterative methods linear systems [96], in finite Markov chain theory [83], input-output analysis in economics, Lotka-Volterra models [90], Google's page rank algorithm, in demographics (Leslie Models) and even in the ranking of American Football teams [56] ...

It is an important theorem and also a little forgotten in the academic courses. The theory of monotonic systems, seen from the angle of dynamical systems, was introduced by Hirsch in a series of papers in the 1980s [43, 45, 46, 47]. In fact they were called cooperative systems. The two terms still coexist. The book "theory of the chemostat" [84] gives some applications of monotonous systems. We have included this theory in this course, because on one hand there are applications in epidemiology and on the other hand there is no elementary references as remarked by H. Thieme.

Chapter 6

Models with continuous delays

6.1 Introduction

Dynamic models of many processes in mathematical epidemiology give systems of ordinary differential equations called compartmental systems. Often, these systems include time lags; in this context, continuous probability density functions (pdfs) of lags are far more important than discrete lags.
Consider the already seen example of an intra-host model

$$\begin{cases} \dot{x} &= \Lambda - \mu_x\, x - \beta\, x\, v \\ \dot{y} &= \beta\, x\, v - \mu_y\, v \\ \dot{m} &= r\, \mu_y\, y - \mu_v\, m - \beta\, x\, v \end{cases} \tag{6.1}$$

variables :

- x Concentration of target cells, namely CD^+ T-cells
- y Concentration of cells invade by virions (HIV)
- m Concentration for free circulating virions

This a classical Model : Perelson (1993), Nowak (1996), May-Anderson-Gupta (1989). In [76] is said fundamental model of immunology.

The binding of a parasite particle to a receptor on a target cell initiates a cascade of events that can ultimately lead to the target cell becoming productively infected, i.e. producing new parasite.

However, in reality there is a time delay between initial viral entry into a cell and subsequent viral production Fixed delays (of which a zero delay is a special case) are not biologically realistic.

For example conversion of a newly infected cell into a productively infected cell is a multi-step process that requires viral entry into the host cell, reverse transcription of viral RNA into DNA, transport of the newly made DNA into the nucleus, integration of the viral DNA into the chromosome, production of viral RNA and protein, and the creation of new virus from these newly synthesized RNA molecules and proteins.

Even in a homogeneous population of target cells, it is unreasonable to expect that the time to complete all of these processes will be the same for every infected cell. If we consider biologically realistic differences in cell activation state, metabolism, position of the cell in the cell cycle, pre-existing stores of nucleotides and other precursors needed for the production of new virions, along with generic variation in the viral population, variation in infection delay times becomes a near certainty. To incorporate a delay, we consider the following modified model

$$\begin{cases} \dot{x} &= \Lambda - \mu_x\, x - \beta\, x\, v \\ \dot{y} &= \beta \displaystyle\int_0^{\infty} g(\tau)\, x(t-\tau)\, v(t-\tau)\, d\tau \; - \mu_y\, y \\ \dot{m} &= r\, \mu_y\, y - \mu_v\, v - \beta\, x\, v \end{cases} \tag{6.2}$$

This is an integrodiff?rential equation. The function $g(t)$ is the probability that a individual infected is infectious t unit of time later. This is a probability density function (pdf). A pdf g satisfies

- $g(t) \geq 0$
- $\displaystyle\int_{-\infty}^{+\infty} g(t)\, dt = 1$
- $F(t) = \text{Prob}[\, X \leq t\,] = \displaystyle\int_{-\infty}^{t} g(s)\, ds.$

A very useful pdf is the Erlang distribution

$$g_{n,\sigma} = \frac{t^{n-1}\, e^{-\frac{t}{\sigma}}}{\sigma^n\, (n-1)!} \tag{6.3}$$

The characteristics are : mean $n\sigma$, deviation $n\sigma^2$, maximum $(n-1)\sigma$; This function is also defined with the parameter $k = \dfrac{1}{\sigma}$. The parameter n is called the shape parameter and σ the rate parameter.

This gamma distribution can reproduce a variety of biological delay distributions and is amenable enough to allow for analytical solutions.

It is a 'tunable' distribution that can mimic both exponential declines and more general bell-shaped distributions

This distribution was developed by Agner Krarup Erlang to model the number of simultaneous phone calls. The Erlang distribution can be used to model the time to complete n operations in series, where each operation requires an exponential period of time to complete. In a Poisson process the sum of n inter-arrival times has an Erlang distribution with parameters n and σ. We will show that the Erlang distribution is the distribution of the sum of k independent and identically distributed random variables each having an exponential distribution..

6.2 Some historical background

When g is a gamma function or a convex combination of gamma distributions the system (6.2) can be converted into a system of differential equations. This have been used in [74]. The process of converting time-delay integro-differential equations in a set of ODE is coined by MacDonald as the "linear chain trick" [73]. In other community this is also known as the method of stages [22, 36, 67, 66, 68]. Any distribution can be approximated by a combination of stages in series and in parallel [52, 22]. Actually it can be proved that given any distribution g with support on $[0, \infty)$, there is a sequence of convex combination of gamma distributions which converges weakly to this distribution. If g is continuous, weak convergence implies uniform convergence on compact intervals. Least squares approximation on a finite interval is used in [54].

In the introduction of his well known monograph on lags [73], MacDonald points out that early in this century, differential equations with lags had been interpreted as equations for systems for which there were hidden state variables.

That idea goes back to Picard who introduced the idea of hidden variables to explain non-conservative systems and then suggested that hereditary effects appear because too few state variables were taken into account

Vogel developed the idea of hidden variables in considerable detail in his monograph in the context of the reduction of hereditary systems to dynamical systems. The use of a catenary chain of compartments with one-way flow to generate a distribution of lags is well known. Gy?ri gives two references in the Russian literature on this chain method and MacDonald used it extensively in his monograph, calling it the Ôlinear chain trickÕ, and pointing to it as an example of hidden variables.

6.3 The Linear Chain Trick

Suppose that the material leaving a compartment j has a pdf of delay $h_{ij}(\tau)$ before entering the compartment i. At time t a fraction $h_{ij}(t-\tau)\,d\tau$ will enter compartment i during the interval $[\tau, t+d\tau]$ for $\tau \leq t$. Then the total material coming from j at time t is

$$f_{ij} \int_{-\infty}^{t} q_j(\tau)\, h_{ij}(t-\tau)\, d\tau$$

To fix the idea consider the following SIR model, where to be infectious has a delay, modeled by a pdf.

$$\begin{cases} \dot{S} &= \Lambda - \beta\, S(t) \displaystyle\int_0^{+\infty} f(\tau)\, I(t-\tau)\, d\tau - \mu\, S(t) \\ \\ \dot{I} &= \beta\, S(t) \displaystyle\int_0^{+\infty} f(\tau)\, I(t-\tau)\, d\tau - (\mu+\gamma)\, I(t) \end{cases}$$

When the pdf is an Erlang function or a convex linear combination of Erlang functions this system is equivalent to a system of ODE [21, 31].
Consider the following input-ouput linear system, which a catenary compartmental system

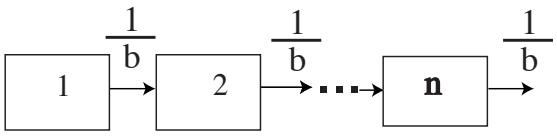


Figure 6.1: Catenary system

The system of ODE is given by

$$\begin{cases} \dot{x}_1 = & u - \frac{1}{\sigma} x_1 \\ \dot{x}_2 = & \frac{1}{\sigma}(x_1 - x_2) \\ \cdots & \\ \dot{x}_n = & \frac{1}{\sigma}(x_{n-1} - x_n) \\ \text{output} & = y = h(x) = \frac{1}{\sigma}\, x_n = \dfrac{1}{\sigma}\, \langle e_n | x \rangle \end{cases}$$

If we denote by e_i the canonical basis of $\mathbb{R}^n_+$, let define $C = \dfrac{1}{\sigma}\, e_n^T$, $B = e_1$ and

$$A = \frac{1}{\sigma} \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}$$

Then the linear system is now

$$\begin{cases} \dot{x} & = A\, x + B\, u \\ y & = C\, x \end{cases}$$

The output is given, for $x_0 = 0$, by

$$y(t) = \int_0^t C\, e^{(t-s)A}\, B\, u(s)\, ds = \int_0^{+\infty} C\, e^{(t-s)A}\, B\, Y(t-s)\, u(s)\, ds$$

where $Y(t)$ is the Heaviside function

$$Y(t) = \begin{cases} 1 & t \geq 0 \\ 0, & t < 0 \end{cases}$$

We recall the convolution of functions :

Definition 6.3.1
If f and g are two locally integrable functions, their convolution, denoted by $f \star g$, is defined by

$$f \star g(t) = \int_{\mathbb{R}} f(t-s)\, g(s)\, ds = \int_{\mathbb{R}} f(t)\, g(t-s)\, ds$$

Then the output $y(t)$ is simply

$$y(t) = [Ce^{tA}B\,Y(t)] \star u(t),$$

and a delay with pdf $h(\tau)$

$$\int_{-\infty}^{t} q_j(\tau)\, h_{ij}(t-\tau)\, d\tau = \int_{\mathbb{R}} q_j(\tau)\, h_{ij}(t-\tau)\, Y(t-\tau)\, d\tau = [h_{ij}\, Y \star q_j]\,(t)$$

We will now compute $Ce^{tA}B$ for the catenary system. First of all we remark that $A = \dfrac{1}{\sigma}\,(-I_n + N)$. The matrix N is a well known nilpotent matrix with 1 on the first sub-diagonal, N^p is the matrix with 1 on the p sub-diagonal. Therefore N^{n-1} is simply the matrix with only 1 in the $(n,1)$ entry and $N^n = 0$.
The second remark is

$$Ce^{tA}B = \frac{1}{\sigma}\,\langle e_n | e^{tA}\, e_1\rangle = \frac{1}{\sigma}\, e^{tA}(n,1)$$

Using the fact that N and I_n are commuting

$$\frac{1}{\sigma}\, e^{tA}(n,1) = e^{-\dfrac{t}{\sigma}}\, \frac{t^{n-1}}{(n-1)!\,\sigma^n} = g_{n,\sigma}$$

Finally

$$y(t) = g_{n,\sigma}(t) \star u(t)$$

This means that introducing such a catenary chain of compartments between two compartments generates a continuous distribution of delays with density function (6.3).
If one introduces a unit impulse, i.e., a Dirac function δ, into the first compartment of the catenary chain at $t = 0$, the density function for time of exit is exactly the Erlang function. We will precise this now.
We consider a linear control system

$$\begin{cases} \dot{x} &= A\,x + B\,u \\ \\ y &= C\,x \end{cases} \tag{6.4}$$

Definition 6.3.2 (Impulse response) *:*
The matrix

$$h(t) = C\, e^{tA}\, B,$$

is called the impulse response of the linear system (6.4)

Proposition 6.3.1 *:*
The impulse response is the output of a linear control system for a nul initial condition, to a Dirac input.

Proof Consider the following sequences of functions

$$\begin{cases} f_n(t) = n & \text{if } 0 \leq t \leq \frac{1}{n} \\ f_n(t) = 0 & \text{otherwise} \end{cases}$$

These functions are called unit pulse ($\int_R f_n(s)ds = 1$)

If we use f_n as an input we obtain

$$\begin{aligned} y_n(t) &= Ce^{tA}\, n \int_0^{\frac{1}{n}} e^{-sA}\, B\, ds \\ &= Ce^{tA}\, n\, \frac{e^{\frac{1}{n}A} - I}{A}\, B \end{aligned}$$

Therefore

$$\lim_{n \to +\infty} n\, \frac{e^{\frac{1}{n}A} - I}{A} = I$$

The limit of y_n is $Ce^{tA}B$.
This result is obtained immediately with the language of distributions, since δ is a unit for the convolution.

$$y(t) = h(t) \star \delta = h(t)$$

The limit of f_n in distributions sens is the Dirac function δ. It can be shown that, for σ fixed, the sequence of Erlang function (6.3) converges to the Dirac function when $n \to +\infty$.

6.4 Generalized linear chain trick

We will show how to generates a convex combination of Erlang functions.

We consider a dynamic system where a peculiar one dimensional feedback $u(x)$ has been distinguished. For example the xv term appearing in the second equation of the system (6.1).

$$\dot{x} = f(x, u(x)) \tag{6.5}$$

The function f is an application from $\mathbb{R}^n \times \mathbb{R}$ into $\mathbb{R}$. The function u is defined from $\mathbb{R}^n$ to $\mathbb{R}$. The functions f and u are supposed to satisfy conditions which

ensure that for any initial state $x(0) = x_0$ the system (6.5) has a unique solution. Usually a system $\dot{x} = f(x, u)$ is called a controlled system.

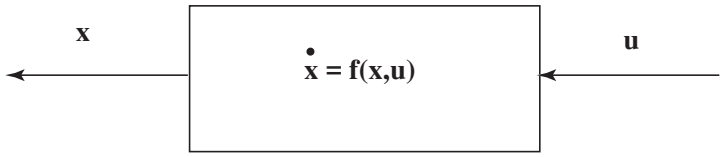


Figure 6.2: Control system

When the function u depends only of time t it is called a control or an input. When u depends on $x(t)$ it is called a feedback.

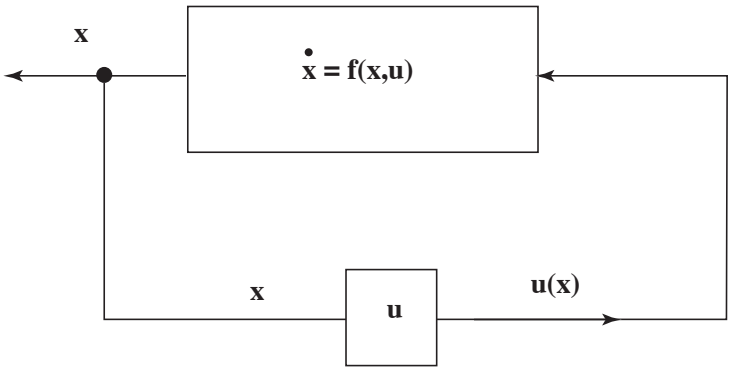


Figure 6.3: Feedback

Let us consider the following controlled linear system (in control theory's sense [87, 71]).

$$\begin{cases} \dot{y} = A\,y + w\,B \\ z = C\,x \end{cases} \tag{6.6}$$

Where the state is $y \in \mathbb{R}^k$, A is a $k \times k$ real matrix, w is a real function, B a $k \times 1$ column vector , $z \in \mathbb{R}$ and C is a $1 \times k$ row vector. In control theory w is the input (or control), z is the output (or observation). We denote by $Y(t)$ the Heaviside function, this function is also known in control theory as the unit step function. For an initial state $y(0) = y_0$, and for a control signal $h(t)$, the output

signal of (5.9) is given by (see for example [63])

$$z(t) = C\,e^{tA}\,y_0 + \int_0^t Ce^{(t-\tau)A}\,B\,h(\tau)\,d\tau \tag{6.7}$$

$$= C\,e^{tA}\,y_0 + \int_{-\infty}^t Ce^{(t-\tau)A}\,B\,h(\tau)\,Y(\tau)\,d\tau \tag{6.8}$$

$$= C\,e^{tA}\,y_0 + \int_0^{+\infty} Ce^{\tau A}\,B\,h(t-\tau)\,d\tau \tag{6.9}$$

$$= C\,e^{tA}\,y_0 + Ce^{tA}B\,Y \star h \tag{6.10}$$

The output is obtained by a convolution integral, where by misuse of language we have denoted by $Ce^{tA}B\,Y(t)$ the impulse response, i.e., the function $t \longmapsto Ce^{tA}B\,Y(t)$. This function is called the impulse response of the system. The reason is that this is the response of the system when the input is the Dirac function considered as a distribution (L. Schwartz's generalized functions, [80]). The output is obtained by convolution of the impulse response with the input. By the classical theory of ODE, $Ce^{tA}B$ is then a linear combination of function of type $t^k\,e^{\lambda t}$, $t^k\,e^{\lambda t}\,\cos(\omega\,t)$ and $t^k\,e^{\lambda t}\,\sin(\omega\,t)$ for $k \in \mathbb{N}$, $\lambda \in \mathbb{R}$ and $\omega \in \mathbb{R}$.

We assume that the kernel function (or the probability density function) h of a certain delay can be represented by $h(t) = Ce^{tA}B\,Y(t)$. The presence of $Y(t)$ is to ensure that the time delay is always positive. Moreover since $h(t)$ is a distribution the matrix A must be a stable matrix. We can assume, without loss of generality that B and C are nonnegative vectors and that the off-diagonal entries of A are nonnegative (which implies e^{tA} is nonnegative). In other words A is a Metzler stable matrix [53, 71].

The system (6.5), when there is a distributed delay (associated to the preceding h) on the feedback, becomes an integro-differential equation.

$$\begin{aligned} \dot{x} &= f(x, \int_0^{\infty} u(x(t-\tau))\,h(\tau)\,d\tau) \\ &= f(x, \int_{-\infty}^{t} u(x(\tau))\,h(t-\tau)\,d\tau) \end{aligned} \tag{6.11}$$

We consider an initial condition

$$u(x(t)) = \theta(t) \text{ for } t \leq 0 \tag{6.12}$$

where θ is a continuous function defined on $]-\infty, 0]$.

Now we define

$$y(t) = \int_{-\infty}^0 e^{(t-\tau)A}\,B\,u(x(\tau))\,d\tau$$

and

$$y_0 = y(0) = \int_{-\infty}^{0} e^{(-\tau)A}\, B\, u(x(\tau))\, d\tau = \int_{-\infty}^{0} e^{(-\tau)A}\, B\, \theta(\tau)\, d\tau$$

We have the relation

$$\int_{-\infty}^{t} u(x(t-\tau))\, h(\tau)\, d\tau = \int_{-\infty}^{t} C\, e^{(t-\tau)A}\, B\, u(x(\tau))\, d\tau = z(t) = C\, y(t)$$

Then for any initial state $x(0) = x_0$ and an initial condition $u(t) = \theta(t)$ on $\mathbb{R}_-$, the integro-differential equation (6.11) is equivalent to

$$\begin{cases} \dot{x} = f(x, Cy) \\ \dot{y} = A\, y + u(x)\, B \end{cases} \tag{6.13}$$

with initial condition $x(0) = x_0 \;\; y(0) = y_0$. More precisely any solution of (6.11) becomes a solution of (6.13) (see [13] for example).
The general linear chain trick can replace system with delays with a pdf which is linear combination of gamma functions, by a system of ODE. Realization theory is a part of linear control theory which gives explicit means of constructing such a matrix A when the distribution is known. The delay is obtained in inserting between the feedback and the original system a linear system.

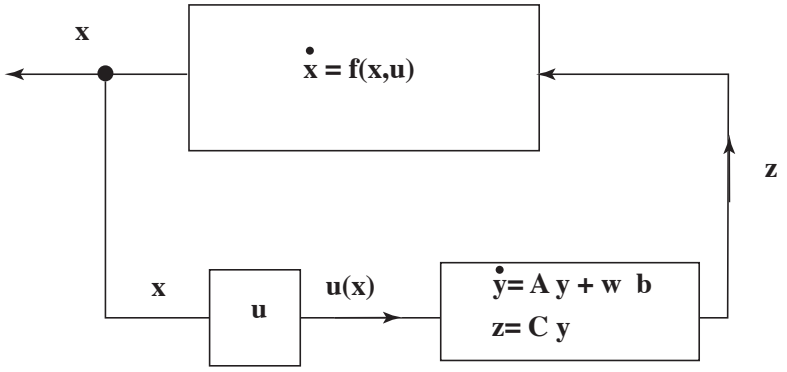


Figure 6.4: Linear chain trick

6.5 Application

When the distribution g is a convex combination of Erlang distribution

$$g = \sum_{i=1}^{q} \pi_i\, g_{k_i,\sigma_i}$$

with $\pi_i \geq 0$ and $\sum \pi_i = 1$, and the corresponding delay is applied to the general class of within-host parasite models (6.1), the system can be replaced by the with the following flow graph (6.5), analogous to figure 9 of [54] represent the block diagram for the interconnection of the different systems :

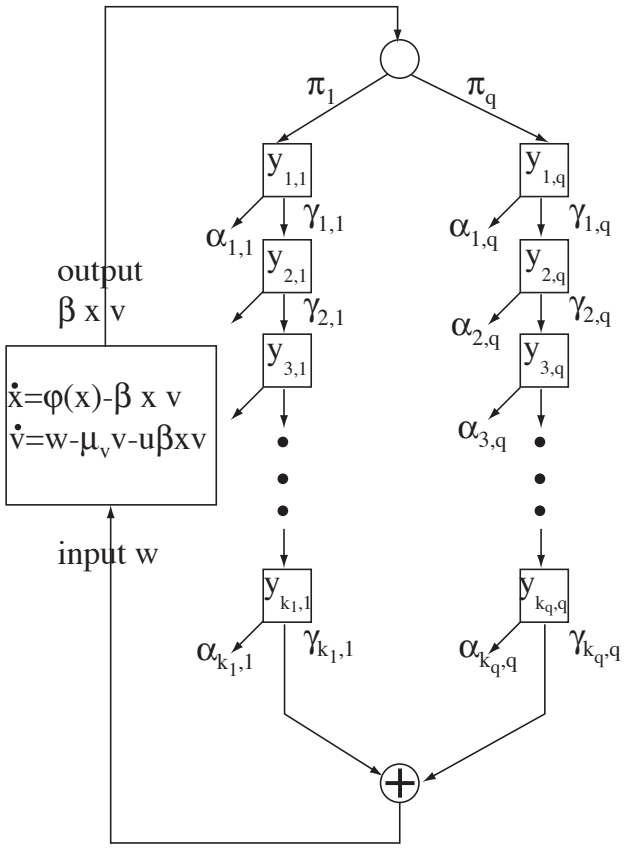


Figure 6.5: Block diagram of system (6.14)

We draw reader's attention to the fact that this block diagram is not strictly speaking a compartmental model, since this is not a mass-balance system. From a pseudo-compartment j is globally leaving a quantity $\alpha_j y_j$ of material, and entering $\gamma_{j-1}\, y_{j-1}$. This $\gamma_{j-1}\, y_{j-1}$ quantity entering in the j compartment has not to be subtracted to the amount of material of the $j-1$ compartment. The coefficients γ_j can be considered as yield coefficients. We only assume that the coefficients are

positive. The arrows are only to symbolize what material is entering (or leaving). This picture is more a signal flow graph in control theory's spirit.
From the flow graph the system of ODE is simply

$$\begin{cases} \dot{x} = \varphi(x) - \beta\, x\, v \\ \\ \text{and for } \ i = 1, \cdots q \\ \\ \dot{y}_{1,i} = \pi_i\, \beta\, x\, v - \alpha_{1,i}\, y_{1,i} \\ \\ \dot{y}_{2,i} = \gamma_{1,i}\, y_{1,i} - \alpha_{2,i}\, y_{2,i} \\ \cdots \\ \dot{y}_{k_i,i} = \gamma_{k_{i-1},i}\, y_{k_{i-1},i} - \alpha_{k_i,i}\, y_{k,i} \\ \\ \dot{v} = \sum_{i=1}^{n}\, \gamma_{k_i,i}\, y_{k_i,i} - \mu_v\, v - u\, \beta\, x\, v \end{cases} \tag{6.14}$$

The system (6.14) can be written in a condensed form.

$$\begin{cases} \dot{x} = \varphi(x) - \beta\, x\, v \\ \dot{y} = A\, y + \beta\, x\, v\, B \\ \dot{v} = C\, y - \mu_v\, v - u\, \beta\, x\, v \end{cases} \tag{6.15}$$

6.5.1 Notations

To simplify the exposition we need some notations. We will adopt some convenient notations from MATLAB or SCILAB. Matrices will be represented by entries between brackets, listed by rows, each element is separated by commas and the semicolon indicates end of the rows.

We denote by $e_k(n)$ the k^{th}-vector of the canonical basis of $\mathbb{R}^n$. In other words for example the vector $e_1(n)$ is the column vector of length n written with our notations $e_1(n) = [1; 0; \cdots; 0]$. We will use the notation $e_{end}(n)$ for the last vector of the canonical basis. We use the same convention to define block matrices, for example $M = [E, F; G, H]$ is the block matrix

$$M = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

provided the matrices E,F,G and H have compatible dimensions. We denote by A^T the transpose of the matrix A. For a vector x of length n we denote by $\mathrm{diag}(x)$ the $n \times n$ diagonal matrix with the elements of x on the diagonal. We also consider $\mathrm{diag}(A_1, \cdots, A_n)$ which is a diagonal block matrix, the A_i being the diagonal blocks.

We can now define A, B and C of (6.15):

The matrix A is a $n \times n$ diagonal block matrix with $n = \sum_{i=1}^{q} k_i$. $A = \text{diag}(A_1, \cdots, A_q)$. The $k_i \times k_i$ block A_i is

$$A_i = \begin{bmatrix} -\alpha_{1,i} & 0 & 0 & \cdots & 0 & 0 \\ \gamma_{1,i} & -\alpha_{2,i} & 0 & \cdots & 0 & 0 \\ 0 & \gamma_{2,i} & -\alpha_{3,i} & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \gamma_{k_i-2,i} & -\alpha_k & 0 \\ 0 & \cdots & 0 & 0 & \gamma_{k_i-1,i} & -\alpha_{k_i,i} \end{bmatrix}$$

The vector B is the column vector of length n

$$B = [\pi_1\, e_1(k_1); \pi_2\, e_1(k_2); \ldots ; \pi_q\, e_1(k_q)]$$

The matrix C is a $1 \times n$ row vector

$$C = [\gamma_{k_1,1}\, e_{end}(k_1)^T, \gamma_{k_2,2}\, e_{end}(k_2)^T, \ldots, \gamma_{k_q,q}\, e_{end}(k_q)^T]$$

The block decompositions of A, B and C are compatible.

6.5.2 Hypotheses

We start to analyze the system with minimal hypothesis on φ but nevertheless plausible from the biological point of view. The function $\varphi(x)$ describes the population dynamics of target cells in absence of parasites. The target cells have a finite lifetime. The function φ models in some way homeostasis. We assume that φ is a $\mathcal{C}^1$ function. Since homeostasis is maintained we assume the system

$$\dot{x} = \varphi(x)$$

has a globally asymptotically stable equilibrium $x^* > 0$, that is,

$$\varphi(x^*) = 0 \quad \varphi(x) > 0 \quad \text{for } 0 \leq x < x^*, \text{ and } \varphi(x) < 0 \text{ for } x > x^*. \tag{6.16}$$

6.6 Stability analysis for the one chain system

To simplify we will examine the system given by a single chain of k elements. We will use the computations of this special case to study the complete system (6.14). In the case of a single chain the system is reduced to

$$\begin{cases} \dot{x} = \varphi(x) - \beta\, x\, v \\ \dot{y} = A\, y + \beta\, x\, v\, B \\ \dot{v} = C\, y - \mu_v\, v - u\, \beta\, x\, v \end{cases} \tag{6.17}$$

with

$$A = \begin{bmatrix} -\alpha_1 & 0 & 0 & \cdots & 0 \\ \gamma_1 & -\alpha_2 & 0 & \cdots & 0 \\ 0 & \gamma_2 & -\alpha_3 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \gamma_{k-1} & -\alpha_k \end{bmatrix}$$

$$C = \gamma_k \, e_{end}(k)^T \text{ and } B = e_1(k)$$

It is clear that the nonnegative orthant is positively invariant by (6.17). The matrix A is a stable Metzler matrix.

6.6.1 Background

For later references we need the expression of the nonnegative matrix $(-A^{-1})$ (A is Metzler stable). Using the fact that $A = -D + N$ where D is the diagonal matrix $D = \text{diag}(\alpha_1, \cdots, \alpha_k)$, N is the nilpotent matrix $N = A + D$. We denote by I the identity matrix, we have $(-A)^{-1} = D^{-1}(I - ND^{-1})^{-1}$. Using the fact that ND^{-1} is nilpotent we get

$$(I - ND^{-1})^{-1} = I + ND^{-1} + (ND^{-1})^2 + \cdots + (ND^{-1})^{k-1}$$

Finally

$$-A^{-1} = \begin{bmatrix} \frac{1}{\alpha_1} & 0 & 0 & \cdots & 0 \\ \frac{\gamma_1}{\alpha_1 \, \alpha_2} & \frac{1}{\alpha_2} & 0 & \cdots & 0 \\ \frac{\gamma_1 \, \gamma_2}{\alpha_1 \, \alpha_2 \, \alpha_3} & \frac{\gamma_2}{\alpha_2 \, \alpha_3} & \frac{1}{\alpha_3} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{\gamma_1 \cdots \gamma_{k-1}}{\alpha_1 \cdots \alpha_k} & \cdots & \cdots & \frac{\gamma_{k-1}}{\alpha_{k-1}\alpha_k} & \frac{1}{\alpha_k} \end{bmatrix}$$

The matrix $-A^{-1}$ is a lower triangular matrix, the i-term on the diagonal is given by $\frac{1}{\alpha_i}$, the entry (i, j) with $i > j$ is

$$(-A^{-1})(i, j) = \frac{\gamma_j \cdots \gamma_{i-1}}{\alpha_j \cdots \alpha_{i-1}} \frac{1}{\alpha_i} \tag{6.18}$$

If we use the usual convention that an empty product has value 1, this expression is also valid for the entries on the diagonal.

6.6.2 Basic reproduction ratio and Equilibria of the system

As usual the basic reproduction number is the expected number of secondary cases produced in a completely susceptible population, by a typical infected individual during its entire period of infectiousness [41, 94, 25, 26]. From the structure of the system the computation of $\mathcal{R}_0$ is straightforward. Indeed one parasite during the mean duration of its life generates a Dirac input $\frac{\beta\, x^*}{\mu_v + u\,\beta\, x^*}$ in the second controlled system $\dot{y} = A\, y + w\, B$. Hence this input generates secondary cases given by the formula

$$\frac{\beta\, x^*}{\mu_v + u\,\beta\, x^*} \int_0^{+\infty} C e^{tA} B\, dt = \frac{\beta\, x^*}{\mu_v + u\,\beta\, x^*} C\,(-A^{-1})\, B$$

This proves

$$\mathcal{R}_0 = \frac{\beta\, x^*}{\mu_v + u\,\beta\, x^*} C\,(-A^{-1})\, B \tag{6.19}$$

With our definition we have $C(-A^{-1})B = \gamma_k\, e_{end}(k)^T(-A^{-1})e_1(k)$ which is simply the entry of the last row , first column of $-A^{-1}$ multiplied by γ_k. Finally

$$\mathcal{R}_0 = \frac{\beta\, x^*}{\mu_v + u\,\beta\, x^*}\, \frac{\gamma_1 \cdots \gamma_k}{\alpha_1 \cdots \alpha_k}. \tag{6.20}$$

We also define a threshold $\mathcal{T}_0$ by

$$\mathcal{T}_0 = \frac{\beta\, \left[\frac{\gamma_1 \cdots \gamma_k}{\alpha_1 \cdots \alpha_k} - u\right] x^*}{\mu_v} \tag{6.21}$$

We call $\mathcal{T}_0$ a threshold since $\mathcal{T}_0 \leq 1$ is equivalent to $\mathcal{R}_0 \leq 1$.

The system has two nonnegative equilibria. The first, called the parasite free equilibrium, is $(x^*, 0 \cdots, 0)$. The second is called the endemic equilibrium and is denoted by $(\bar{x}, \bar{y}, \bar{v})$.
We have necessarily $\bar{y} = \beta\, \bar{x}\, \bar{v}(-A^{-1})\, e_1$ and

$$\mu_v\, \bar{v} + u\,\beta\, \bar{x}\, \bar{v} = \gamma_k\, \beta\, \bar{x}\, \bar{v}\, e_{end}^T(-A^{-1})\, e_1$$

If $\bar{v} \neq 0$ we deduce

$$\bar{x} = \frac{\mu_v}{\beta\, \left[\gamma_k\, e_{end}^T\,(-A^{-1})\, e_1\, - u\right]} = \frac{x^*}{\mathcal{T}_0}$$

With this expression we get $\bar{v} = \frac{\varphi(\bar{x})}{\beta\, \bar{x}}$. Hence, with the hypothesis (6.16), $\bar{x}$ and $\bar{v}$ are positive iff $\mathcal{T}_0 > 1$ or equivalently iff $\mathcal{R}_0 > 1$. Now $\bar{y} = \varphi(\bar{x})(-A^{-1})\, e_1$. In other words $\bar{y}$ is the first column of $(-A^{-1})$ multiplied by $\varphi(\bar{x})$. The first column of $(-A^{-1})$ is a positive vector, hence v is in the positive orthant, classically denoted by $\bar{v} \gg 0$.

To summarize the endemic equilibrium is in the positive orthant iff $\mathcal{R}_0 > 1$ and it is given by

$$\begin{cases} \bar{x} = \frac{\mu_v}{\beta\left[\frac{\gamma_1\cdots\gamma_k}{\alpha_1\cdots\alpha_k} - u\right]} = \frac{\bar{x}}{\mathcal{I}_0} \ < \ x^* \\ \bar{y} = \varphi(\bar{x})\,(-A)^{-1}e_1 \\ \bar{v} = \frac{\varphi(\bar{x})}{\beta\,\bar{x}} \end{cases} \tag{6.22}$$

6.6.3 Stability analysis

We give the main result of the section

Theorem 6.6.1 *We consider the system (6.15) with the hypothesis on φ (6.16) satisfied. The basic reproduction ratio of the system is given by (6.20).*

1. *The system (6.15) is globally asymptotically stable on $\mathbb{R}^{k+2}_+$ at the parasite free equilibrium (PFE) $(x^*, 0, \cdots, 0)$ if and only if $\mathcal{R}_0 \leq 1$.*
2. *If $\mathcal{R}_0 > 1$ then the PFE is unstable and there exists a unique endemic equilibrium (EE) in the positive orthant, $(\bar{x}, \bar{y}, \bar{v}) \gg 0$ given by (6.22)*
3. *If $\mathcal{R}_0 > 1$, denoting $\alpha^* = -\max_{x\in[0,x^*]}(\varphi'(x))$, and if*

$$u\,\beta\,\varphi(\bar{x}) \leq \alpha^*\,\mu_v \tag{6.23}$$

then the endemic equilibrium is globally asymptotically stable on the nonnegative orthant, excepted for initial conditions on the x-axis.

Remark 6.6.1 *If φ increases on some part of its domain, the relation (6.23) is never satisfied. In this happens, it may lead to limit cycle for this model as in [23].*

Remark 6.6.2 *When $u = 0$ and $\varphi(x) = \Lambda - \mu_x\,x$ the sufficient condition is automatically satisfied. This is the case of numerous models of the literature. See for example the general model (1) of [75] or the model in [67].*

Proof. We need some dissipativity properties of system (6.15). In a first step we show that there exists in the nonnegative orthant $\mathbb{R}^{k+2}_+$ a forward invariant compact absorbing neighborhood Ω of the PFE $(x^*, 0, \cdots, 0)$. An absorbing set D is a neighborhood of the PFE such that the trajectory of the system starting from any initial condition enters and remains in D for a sufficiently large time T. The entrance time depends on the initial condition. If the initial conditions are contained in a compact set F then there exits a uniform T for F. A system is point dissipative if there exists a compact absorbing set. The above definition coincides with dissipativity given by [19].

Let $\varepsilon \geq 0$ be a given nonnegative real. With the hypothesis (6.16) on φ there exists a time T such that, for any initial condition in the nonnegative orthant and for $t \geq T$ we have $x(t) \leq x^* + \varepsilon$. Let M_φ be the maximum of the function $\varphi(x)$ on $\mathbb{R}_+$. Let A a positive real such that $\alpha_1 A > M_\varphi + \varepsilon$.

We claim that the set $\mathcal{D}_\varepsilon$ defined by

$$\mathcal{D}_\varepsilon = \Big\{(x,y,v) \in \mathbb{R}_+^{k+2} \mid x \leq x^* + \varepsilon,\ x + y_1 \leq A + x^* + \varepsilon, \quad \text{and for } i = 2 \cdots k \\ y_i \leq \tfrac{\gamma_2 \cdots \gamma_{i-1}}{\alpha_2 \cdots \alpha_i}(A + x^* + i\varepsilon), v \leq \tfrac{\gamma_1 \cdots \gamma_k}{\alpha_2 \cdots \alpha_k\, \mu_v}(A + x^* + k\varepsilon)\ \}$$

is a forward invariant compact absorbing set for the system for $\varepsilon > 0$, and that the set $\mathcal{D}_0$ ($\varepsilon = 0$) is a forward invariant compact set.

The set $\mathcal{D}_\varepsilon$ is the intersection of halfspaces defined by some hyperplanes. To prove the positive invariance of a set, it is sufficient to prove that the vector field associated to the system is tangent or pointing to the set on the boundary of this set. See Theorem (3.1.1). This is immediate for the faces of the nonnegative orthant and for the halfspace defined by $\mathcal{D}_{\varepsilon,1} = \{(x,y,v) \mid x \leq x^* + \varepsilon\}$. From the properties of φ this set is also clearly absorbing. We define $\mathcal{D}_{\varepsilon,2} = \{(x,y,v) \in \mathcal{D}_{\varepsilon,1} \mid x + y_1 \leq A + x^* + 2\,\varepsilon\}$. We have just to look at the boundary of $\mathcal{D}_{\varepsilon,2}$ contained in $\mathcal{D}_{\varepsilon,1}$. On this part of the boundary we have $y_1 \geq A$. Then on this boundary we have $\dot{x} + \dot{y}_1 \leq M_\varphi - \alpha_1\, y_1 < \varepsilon$. The vector is re-entrant, hence $\mathcal{D}_{\varepsilon,2}$ is positively invariant. The inequality $\dot{x} + \dot{y}_1 < \varepsilon$ proves that $\mathcal{D}_{\varepsilon,2}$ is absorbing in $\mathcal{D}_{\varepsilon,1}$. A finite induction process, with similar arguments, ends the proof for $\mathcal{D}_\varepsilon$.

In a second step we will prove that, if $\mathcal{R}_0 \leq 1$, the PFE is globally asymptotically stable on the compact forward invariant set $\mathcal{D}_0$. It is well known that if $\mathcal{R}_0 > 1$ then the PFE is unstable [26, 94] . Thus the condition $\mathcal{R}_0 \leq 1$ is necessary.

To prove the sufficiency we consider the following Liapunov function on the positive orthant.

$$V_{PFE}(y,v) = b^T\, y + v \tag{6.24}$$

where the column vector $b = [b_1; b_2; \cdots; b_k]$ is the transpose of the last row of $-A^{-1}$ multiplied by γ_k. In other words $b = \gamma_k\, (-A^{-T})\, e_{end}$.

We also define for further reference $a = b_1 - u$. If we use (6.22) we obtain for the parameter a the equivalent relation

$$a = b_1 - u = \frac{\mu_v}{\beta\, \bar{x}} = \left[\frac{\gamma_1 \cdots \gamma_k}{\alpha_1 \cdots \alpha_k} - u\right] \tag{6.25}$$

If we compute the derivative of V_{PFE} along the trajectories of (6.15) we get

$$\begin{aligned}\dot{V}_{PFE} &= b^T\,\dot{y} + \dot{v}\\ &= \gamma_k\, e_{end}^T\,(-A^{-1})\,A\,y + \gamma_k\, e_{end}^T\,(-A^{-1})\,\beta\,x\,v\,e_1 + \dot{v}\\ &= -\gamma_k\, y_k + \beta\,x\,v\,b_1 + \gamma_k\, y_k - \mu_v\,v - u\beta\,x\,v\\ &= v[(b_1-u)\beta x - \mu_v] = v[a\,\beta\,x - \mu_v]\\ &= v[\tfrac{\mu_v\,x}{\bar{x}} - \mu_v]\\ &= \tfrac{\mu_v}{\bar{x}}(x-\bar{x})v\\ &= \beta\,\left[\tfrac{\gamma_1\cdots\gamma_k}{\alpha_1\cdots\alpha_k} - u\right](x-\bar{x})v\end{aligned}$$

If $\mathcal{R}_0 \le 1$, or equivalently $\mathcal{T}_0 \le 1$, we distinguish two cases :

1. On one hand if $(\frac{\gamma_1\cdots\gamma_k}{\alpha_1\cdots\alpha_k} - u) < 0$ then $\bar{x} < 0$ and all the other quantities are nonnegative in the expression of $\dot{V}$. Therefore $\dot{V} \le 0$.
2. On the other hand if $(\frac{\gamma_1\cdots\gamma_k}{\alpha_1\cdots\alpha_k} - u) \ge 0$, then from $\mathcal{T}_0 \le 1$ and since we are in $\mathcal{D}_0$ we deduce $0 \le x \le x^* \le \bar{x}$ it follows that $\dot{V} \le 0$.

In both cases $\dot{V} \le 0$. It is easy to see that the maximum invariant set in $\{(x,y,v) \in \mathcal{D}_0 \mid \dot{V} = 0\}$ is reduced to the PFE. Therefore the global asymptotic stability of the PFE on the compact positively invariant set $\mathcal{D}_0$ follows from ([15], Theorem 3.7.11, page 346). Now, We will prove the global asymptotic stability on the orthant $\mathbb{R}_+^{k+2}$. It is sufficient to prove that any forward trajectory converges to the PFE. Since $\mathcal{D}_1$ (i.e $\mathcal{D}_\varepsilon$ for $\varepsilon = 1$) is a forward compact absorbing set any trajectory enters $\mathcal{D}_1$. If a trajectory enters the interior of $\mathcal{D}_0$ we have already proved that it converges toward the PFE. Now assume that a trajectory, in $\mathcal{D}_1$ stays in $\mathcal{D}_1 \cap \{x^* \le x \le x^* + 1\}$. Consider the Liapunov function $W(x) = \frac{1}{2}(x - x^*)^2$ on this trajectory. By the hypothesis (6.16) on φ and the hypothesis on the trajectory we have $\dot{W} = (x-x^*)\varphi(x) - (x-x^*)\,\beta x\,v \le 0$ on any point of the trajectory in $\mathcal{D}_1$. By LaSalle's principle it follows from $\dot{W} \le 0$ that the PFE is the largest invariant set contained in $\{x \in D_1, x^* \le x \le x^* + 1 \mid \dot{W} = 0\}$. This ends the proof of the GAS of the PFE.

Now we assume that $\mathcal{R}_0 > 1$. The equilibria $(\bar{x}, \bar{y}, \bar{v})$ of the system, different from the PFE, is given by (6.22) is in the positive orthant since $\mathcal{R}_0 > 1$.

We will now prove a sufficient condition for the GAS of the EE. To this end we define the following Liapunov function on the positive orthant.

$$V_{EE}(x,y,v) = a(x - \bar{x}\ln x) + \sum_{i=1}^{k} b_i\,(y_i - \bar{y}_i \ln y_i) \;+\; (v - \bar{v}\,\ln v) \tag{6.26}$$

where the column vector b and the coefficient a have been previously defined by the relation (6.25). Since $\mathcal{R}_0 > 1$ we deduce $a > 0$, hence the coefficients of V_{EE}

are positive. In this case this function has a unique minimum, the EE, in the positive orthant.

This function has a linear part $L_{EE}(x,y,v) = a\,x + \sum_{i=1}^{k} b_i\,y_i + v$. This linear part can be expressed as

$$L_{EE}(x,y,v) = a\,x + b^T\,y + v = a\,x + \gamma_k\,e_{end}\,(-A^{-1})\,y + v$$

If we compute the derivative $\dot{L}_{EE}$ of L_{EE} along the trajectories of (6.15), considering the definition of b and the relation $a + u = b_1$, we get

$$\begin{aligned} L_{EE}(x,y,v) &= a\,\dot{x} + \gamma_k\,e_{end}\,(-A^{-1})\,\dot{y} + \dot{v} \\ &= a\,\dot{x} + \gamma_k\,e_{end}\,(-A^{-1})\,A\,y + \beta\,x\,v\,e_{end}\,(-A^{-1})B + \dot{v} \\ &= a\,\dot{x} - \gamma_k\,e_{end}\,y + \beta\,x\,v\,e_{end}\,(-A^{-1})\,e_1 + \dot{v} \\ &= a\varphi(x) - a\beta xv - \gamma_k\,y_k + b_1\,\beta\,x\,v\; + \gamma_k\,y_k - \mu_v\,v - u\,\beta\,x\,v \\ &= a\varphi(x) - \mu_v\,v \end{aligned}$$

If we collect in $\dot{V}_{EE}$ the terms in v we obtain $(a\beta\bar{x} - \mu_v)v$. From (6.25) the terms in v cancel . With these simplifications we can now express $\dot{V}_{EE}$

$$\begin{aligned} \dot{V}_{EE} \;\dot{=}\; & a\,\varphi(x)\left(1 - \tfrac{\bar{x}}{x}\right) - b_1\,\beta\bar{y}_1\tfrac{xv}{y_1} - \textstyle\sum_{i=2}^{k} b_i\gamma_{i-1}y_{i-1}\tfrac{\bar{y}_i}{y_i} + \\ & + \sum_{i=1}^{k} b_i\alpha_i\bar{y}_i - \gamma_k y_k\frac{\bar{v}}{v} + u\,\beta\bar{v}x + \mu_v\bar{v}. \end{aligned}$$

This can also be written

$$\begin{aligned} \dot{V}_{EE} &= a\,\varphi(x)\left(1 - \frac{\bar{x}}{x}\right) - b_1\,\beta\bar{x}\,\bar{v}\frac{x}{\bar{x}}\,\frac{v}{\bar{v}}\,\frac{\bar{y}_1}{y_1} - \sum_{i=2}^{k} b_i\,\gamma_{i-1}\,\bar{y}_{i-1}\frac{y_{i-1}}{\bar{y}_{i-1}}\,\frac{\bar{y}_i}{y_i} + \\ &+ \sum_{i=1}^{k} b_i\alpha_i\bar{y}_i - \gamma_k\,\bar{y}_k\frac{y_k}{\bar{y}_k}\,\frac{\bar{v}}{v} + u\,\beta\bar{x}\,\bar{v}\frac{x}{\bar{x}} + \mu_v\bar{v} \end{aligned}$$

We now compare some coefficients appearing in this formula. We have $\varphi(\bar{x}) = \beta\,\bar{x}\bar{v}$. Using the fact that $\bar{y}$ is the first column of $-A^{-1}$ multiplied by $\varphi(\bar{x})$, b is the transpose of the last row multiplied by γ_k of the same matrix and accordingly to the relation (6.18), we can now consider $b_i\,\gamma_{i-1}\bar{y}_{i-1}$:

$$\begin{aligned} b_i\,\gamma_{i-1}\bar{y}_{i-1} &= \gamma_k\,(-A^{-1})(k,i)\,\gamma_{i-1}\,\varphi(\bar{x})\,(-A^{-1})(i-1,1) \\ &= \varphi(\bar{x})\,\gamma_k\,\frac{\gamma_i\cdots\gamma_{k-1}}{\alpha_i\ldots\alpha_{k-1}}\frac{1}{\alpha_k}\,\gamma_{i-1}\,\frac{\gamma_1\cdots\gamma_{i-2}}{\alpha_1\cdots\alpha_{i-2}}\,\frac{1}{\alpha_{i-1}} \\ &= \frac{\gamma_1\cdots\gamma_k}{\alpha_1\cdots\gamma_k}\,\varphi(\bar{x}) = b_1\,\varphi(\bar{x}) = b_1\,\beta\,\bar{x}\bar{v} \end{aligned}$$

In the same way

$$\begin{aligned} b_i\,\alpha_i\bar{y}_i &= \gamma_k\,(-A^{-1})(k,i)\,\alpha_i\,\varphi(\bar{x})\,(-A^{-1})(i,1) \\ &= \varphi(\bar{x})\,\gamma_k\,\frac{\gamma_i\cdots\gamma_{k-1}}{\alpha_i\ldots\alpha_{k-1}}\frac{1}{\alpha_k}\,\alpha_i\,\frac{\gamma_1\cdots\gamma_{i-1}}{\alpha_1\cdots\alpha_{i-1}}\,\frac{1}{\alpha_i} = \frac{\gamma_1\cdots\gamma_k}{\alpha_1\cdots\gamma_k}\,\varphi(\bar{x}) = b_1\,\varphi(\bar{x}) \end{aligned}$$

and $\gamma_k\,\bar{y}_k = \gamma_k\,(-A^{-1})(k,i)\,\varphi(\bar{x}) = \varphi(\bar{x})\,\gamma_k\,\dfrac{\gamma_1\cdots\gamma_{k-1}}{\alpha_1\ldots\alpha_{k-1}}\dfrac{1}{\alpha_k} = b_1\,\varphi(\bar{x})$.

According to (6.25) we also have $\mu_v\bar{v} = a\,\beta\,\bar{x}\,\bar{v} = a\,\varphi(\bar{x})$

Using all these relations between the coefficients we get for $\dot{V}_{EE}$

$$\begin{aligned} \dot{V}_{EE} &= a\,\varphi(x)\left(1-\frac{\bar{x}}{x}\right)+ \\ &\quad b_1\,\varphi(\bar{x})\left[k-\frac{x}{\bar{x}}\,\frac{v}{\bar{v}}\,\frac{\bar{y}_1}{y_1}-\sum_{i=2}^{k}\frac{y_{i-1}}{\bar{y}_{i-1}}\,\frac{\bar{y}_i}{y_i}-\frac{y_k}{\bar{y}_k}\,\frac{\bar{v}}{v}\right]+u\,\varphi(\bar{x})\frac{x}{\bar{x}}+a\,\varphi(\bar{x}) \end{aligned}$$

Adding $2-\frac{\bar{x}}{x}$ in the expression between brackets and subtracting the same expression outside the brackets, using $u = b_1 - a$ we obtain

$$\begin{aligned} \dot{V}_{EE} &= a\,\varphi(x)\left(1-\frac{\bar{x}}{x}\right)+ \\ &\quad b_1\,\varphi(\bar{x})\left[k+2-\frac{\bar{x}}{x}-\frac{x}{\bar{x}}\,\frac{v}{\bar{v}}\,\frac{\bar{y}_1}{y_1}-\sum_{i=2}^{k}\frac{y_{i-1}}{\bar{y}_{i-1}}\,\frac{\bar{y}_i}{y_i}-\frac{y_k}{\bar{y}_k}\,\frac{\bar{v}}{v}\right] \\ &\quad +b_1\,\varphi(\bar{x})\left(\frac{\bar{x}}{x}+\frac{x}{\bar{x}}-2\right)+a\,\varphi(\bar{x})\left(1-\frac{x}{\bar{x}}\right) \end{aligned}$$

If we factor, in this expression, $\frac{x-\bar{x}}{x\,\bar{x}}$ we obtain

$$\begin{aligned} \dot{V}_{EE} &= \frac{x-\bar{x}}{x\bar{x}}\,(a\,\bar{x}\,\varphi(x)-a\,x\,\varphi(\bar{x})+b_1\,\varphi(\bar{x})(x-\bar{x}))+ \\ &\quad b_1\,\varphi(\bar{x})\left[k+2-\frac{\bar{x}}{x}-\frac{x}{\bar{x}}\,\frac{v}{\bar{v}}\,\frac{\bar{y}_1}{y_1}-\sum_{i=2}^{k}\frac{y_{i-1}}{\bar{y}_{i-1}}\,\frac{\bar{y}_i}{y_i}-\frac{y_k}{\bar{y}_k}\,\frac{\bar{v}}{v}\right] \end{aligned}$$

Now we will use the fact that there exists ξ in the open interval $]x,\bar{x}[$ such that $\varphi(x) = \varphi(\bar{x}) + (x-\bar{x})\,\varphi'(\xi)$. Replacing in the preceding expression gives

$$\begin{aligned} \dot{V}_{EE} &= \frac{(x-\bar{x})^2}{x\bar{x}}\left(-a\,\varphi(\bar{x})+a\,\bar{x}\,\varphi'(\xi)+b_1\,\varphi(\bar{x})\right)+ \\ &\quad b_1\,\varphi(\bar{x})\left[k+2-\frac{\bar{x}}{x}-\frac{x}{\bar{x}}\,\frac{v}{\bar{v}}\,\frac{\bar{y}_1}{y_1}-\sum_{i=2}^{k}\frac{y_{i-1}}{\bar{y}_{i-1}}\,\frac{\bar{y}_i}{y_i}-\frac{y_k}{\bar{y}_k}\,\frac{\bar{v}}{v}\right] \end{aligned}$$

Finally since $a - b_1 = u$

$$\begin{aligned} \dot{V}_{EE} &= \frac{(x-\bar{x})^2}{x\bar{x}}\left(u\,\varphi(\bar{x})+a\,\bar{x}\,\varphi'(\xi)\right)+ \\ &\quad b_1\,\varphi(\bar{x})\left[k+2-\frac{\bar{x}}{x}-\frac{x}{\bar{x}}\,\frac{v}{\bar{v}}\,\frac{\bar{y}_1}{y_1}-\sum_{i=2}^{k}\frac{y_{i-1}}{\bar{y}_{i-1}}\,\frac{\bar{y}_i}{y_i}-\frac{y_k}{\bar{y}_k}\,\frac{\bar{v}}{v}\right] \end{aligned}$$

The term between brackets in the last expression of $\dot{V}$ is non positive by the inequality between the arithmetical mean and the geometrical mean. Therefore a sufficient condition for $\dot{V} \leq 0$ is

$$u\,\varphi(\bar{x}) + a\,\bar{x}\,\varphi'(\xi) \leq 0$$

Or equivalently since $a\bar{x} = \frac{\mu_v}{\beta}$

$$u\,\beta\,\varphi(\bar{x}) \leq -\varphi'(\xi)\,\mu_v$$

Moreover with this condition $\dot{V}$ is negative excepted at the EE for the system (6.15). If

Since V_{EE} is a proper function on the positive orthant, this proves the GAS of the EE on the positive orthant for the system (6.15).

The vector field associated with the system is strictly entrant on the faces of the orthant except on the x-axis where it is tangent. The basin of attraction of the EE is then the orthant excepted the one-dimensional face contained in the x-axis of the orthant, which is the stable manifold of the PFE.

Setting $\alpha^* = -\max_{x\in[0,x^*]} \varphi'(x)$, a sufficient condition for the GAS of the EE is

$$\mathcal{R}_0 > 1 \quad \text{and} \quad u\,\beta\,\varphi(\bar{x}) \leq \mu_v \alpha^*$$

This ends the proof of the theorem.

■

6.7 Stability for the complete system

The proof use the same line of ideas. This is proven in [50].

Chapter 7

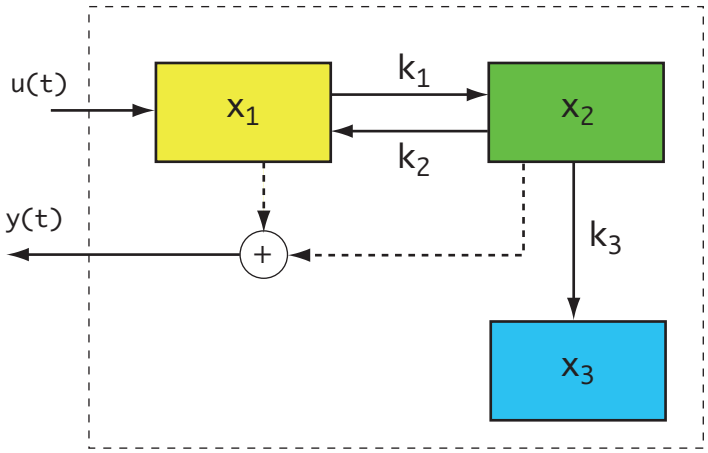
Identification of parameters.

7.1 Introduction

ODE has been widely used to model epidemics of infectious diseases or the intra-host dynamics of a parasite. The analysis of such models, their simulation has attracted a great attention and are the topics of many research papers. This is the problem of predicting the result of measurement, modelization, simulation problems. However, less effort has been devoted to the so-called inverse problem. The inverse problem consists of using the actual results of some measurement to infer the values of the parameters that characterize the system.
This problem, depending of the scientific community, has also received different names : data assimilation, estimation of parameters or identification. In this chapter we will use the concept emanating from control theory of identification of parameters.
Actually, before parameter estimation algorithms can be used to an ODE model to estimate the model parameters based measurements, a serious lock must be overcome : how to verify whether the model parameters are identifiable based on the measurements of output variables. In other words, with the knowledge of some measurements, is there a unique set of parameters which gives these measures ? Does the inverse problem have a unique solution ?
This is the problem of identifiability : whether or not is it possible to distinguish different sets of parameters from the measurement of the output. Before the introduction of precise definitions we will provide a simple example.

7.1.1 A non identifiable linear system

We consider the following system, whose flow graph is



$$\begin{cases} \dot{x}_1 = & -k_1\, x_1 + k_2\, x_2 + u(t) \\ \dot{x}_2 = & k_1\, x_1 - (k_2 + k_3)\, x_2 \\ \dot{x}_3 = & k_3\, x_2 \\ y = & x_1 + x_2 \end{cases}$$

The only known quantities are the observation $y(t)$ and the input $u(t)$. Since y and u are known, $\dot{y} = -k_3\, x_2 + u$ is also known.
Hence

$$\begin{aligned} \ddot{y} &= k_3\, k_1\, x_1 - k_3\, (k_2 + k_3)\, x_2 + \dot{u} \\ &= k_3\, k_1\, (y - x_2) - k_3\, (k_2 + k_3)\, x_2 + \dot{u} \\ &= k_1\, k_3\, y + (k_1 + k_2 + k_3)\, (\dot{y} - u) + \dot{u} \end{aligned}$$

By induction

$$y(t) = y(0) + t\, \dot{y}(0) +$$
$$\sum_{n \geq 2} \frac{t^n}{n!} \left[k_1\, k_3\, y^{(n-2)}(0) + (k_1 + k_2 + k_3)\, y^{(n-1)}(0) + (k_1 + k_2 + k_3)\, u^{(n-2)}(0) + u^{(n-1)}(0) \right]$$

This means that for any set of parameters satisfying

$$(\widetilde{k_1} + \widetilde{k_2} + \widetilde{k_3}) = (k_1 + k_2 + k_3) \text{ and } \widetilde{k_1}\,\widetilde{k_3} = k_1\,k_3$$

will gives the same input-output behavior. The system is not identifiable with the output.

7.1.2 Historical Background

For system of ODE it is Kalman [55] who, for the first time mentioned, in a rather cryptic way, the notion of identifiability.
The paper of Belmann and Åström [12] in 1970 gives precise definitions of structural identifiability for linear control systems. The first systematic treatment in 1987 for nonlinear system is by Tunali and Tarn [93].
Identifiability is simply, as we have try in the preceding example, to express parameters as functions of the known quantities of the system, such as input and output. In this aspect, an algebraic definition, its relationship to observability, and algorithmic procedures based on differential algebraic polynomial systems were rigorously studied in [28, 32, 65].

7.2 Concepts from control theory

In this section we will consider the following system

$$\begin{cases} \dot{x} = & X(x, \theta, u) = X^{u,\theta}(x) \\ \\ y = & h(x) \end{cases} \tag{7.1}$$

We assume that the application $X : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^k \longrightarrow \mathbb{R}^n$ is $\mathcal{C}^\infty$ or analytic as $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. The variable x is the state of the system in the state space $\mathbb{R}^n$, θ is a parameter in $\mathbb{R}^p$, and $u \in \mathbb{R}^k$ is a control. The function h is the observation or output of the system. For any initial state x_0 we denote by $X_t^{u,\theta}(x_0)$ the solution at time t and by $y(x_0, u, t) = h(X_t^{u,\theta}(x_0))$ the corresponding output.

7.2.1 Observability

Definition 7.2.1 (indistinguishability) *Two states x_1 and x_2 are said indistinguishable for system (7.1) iff for any time $t \geq 0$ and any input $u(t)$*

$$y(x_1, u(), t) = y(x_2, u(), t)$$

Indistinguishability $\mathcal{I}$ is an equivalence relation on $\mathbb{R}^n$.
Two states x_1, x_2, $x_1 \neq x_2$ are said to be distinguishable if there exists an admissible control (or input) u and a time $t \geq 0$ such that $y(x_1, u, t) \neq y(x_2, u, t)$. An admissible input which distinguishes every pair of states is called an universal input.
Roughly speaking, this means that the information provided by the measurable output is not enough to tell us if the evolution of the system is given by the solution emanating from the state x_1 or by the one emanating from the state x_2.

Definition 7.2.2 (Observability)
A system is said observable if any pair of distinct states (x_1, x_2) *are distinguishable.*

The system (7.1) is observable if $\mathcal{I}(x) = x$ on $\mathbb{R}^n$

Consider the following model of a chemostat

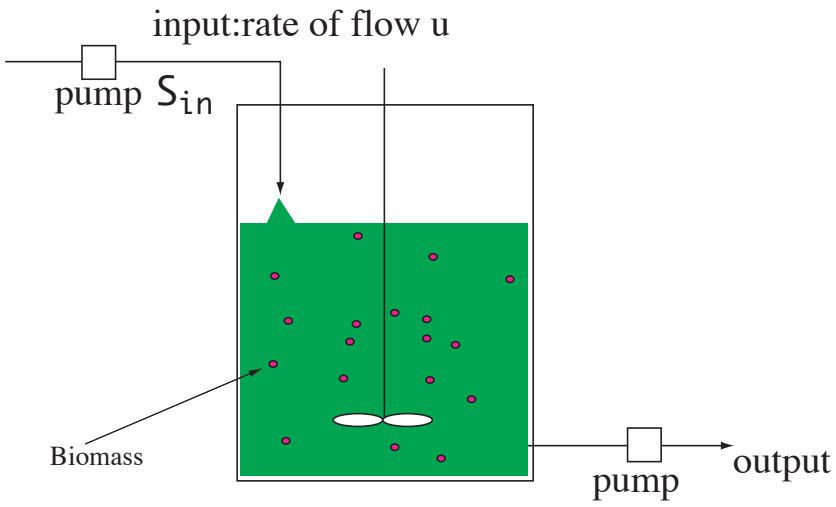


Figure 7.1: Chemostat

$$\begin{cases} \dot{x} = & \mu(s)\, x - u\, x \\ \dot{s} = & -k\, \mu(s) - u\, (s - s_{\text{in}}) \\ y = & x \end{cases}$$

where $x(t)$ and $s(t)$ are respectively the concentration in micro-organisms and substrate, the function $\mu(s)$ is the absorbing rate of the substrate by the micro-organisms and k is the yield coefficient. For this system, the input is the flow rate u and the output is usually the concentration $x(t)$. The substrate is introduced as $u\, s_{\text{in}}$.
There are numerous models for the function μ. For example the Haldane function

$$\mu(s) = \frac{\mu_0\, s}{K_m + s + \frac{s^2}{K_I}}$$

This function is used to model more realistic cases where an excessive concentration of substrate can impede the growth of the biomass. The Haldane function is not injective.

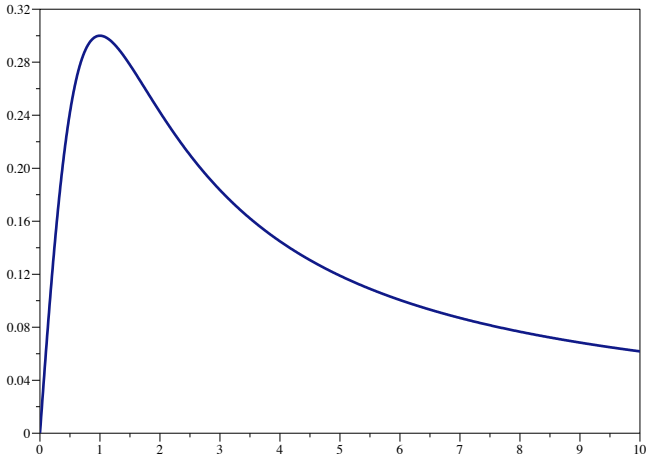


Figure 7.2: Haldane function

Therefore the system is not observable, even if x, u, k and s_{in} are known.
For analytic systems there is a criteria for observability. For giving this criterium we recall some definitions. A smooth vector field operates on $\mathcal{C}^\infty(\mathbb{R}^n$, the set of functions $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$, by Lie differentiation in the following way

$$\begin{array}{ccc} \mathcal{C}^\infty(\mathbb{R}^n) & \longrightarrow & \mathcal{C}^\infty(\mathbb{R}^n) \\ \Phi & \longrightarrow & X.\Phi \end{array}$$

with

$$X.\Phi\,(x) = \frac{d}{dt}\left(\Phi(X_t(x))\right)\Big|_{t=0}$$

As usual $X_t(x)$ denotes the flow of X initiating from x at time t.

The function $X.\Phi$ is called the Lie derivative of Φ along the vector field X. We have the following relation

$$X.\Phi\,(x) = \langle \nabla\Phi(x) \mid X(x)\rangle$$

The Lie derivative of order k is defined by induction

$$X^k.\Phi = X.(X^{k-1}.\Phi)$$

Definition 7.2.3 (Observation space) *The observation space $\mathcal{O}$ of system (7.1) is the linear subspace over $\mathbb{R}$ of functions of $\mathcal{C}^\infty(\mathbb{R}^n)$, containing the observation function h and closed under the Lie differentiation by all elements of*

$$\mathcal{X} = \{f(.,\theta,u) | u \in \mathbb{R}^p\}.$$

We have the description of $\mathcal{O}$

$$\mathcal{O} = span_{\mathbb{R}}\left\{(X^{u_l,\theta})^{k_l}.\ldots.(X^{u_1,\theta})^{k_1}.h \mid l \geq 0,\ u_1,\ldots,u_l \in U,\ k_i \in \mathbb{N}\right\}$$

Theorem 7.2.1 *For analytic systems the observability is equivalent to the fact that the observation space separates the points of $\mathbb{R}^n$, i.e., if $x_1 \neq x_2$ there exists $g \in \mathcal{O}$ such that $g(x_1) \neq g(x_2)$.*

Proof

By analyticity for one constant control u

$$y(x_0,u,t) = h(X_t^{u,\theta}(x_0)) = \sum_{n\geq 0} \frac{t^n}{n!}\frac{d^n}{dt^n}h(X_t^{u,\theta}(x_0))\Big|_{t=0}$$

But

$$\frac{d}{dt}h(X_t^{u,\theta}(x_0))\Big|_{t=0} = \langle \nabla h(x_0) | X^{u,\theta}(x_0)\rangle = X^{u,\theta}.h\,(x_0)$$

By induction

$$\frac{d^n}{dt^n}h(X_t^{u,\theta}(x_0))\Big|_{t=0} = (X^{u,\theta})^n.h\,(x_0)$$

Then it is necessary and sufficient for two distincts initial states $x_1 \neq x_2$, for having different outputs, that there exists an index n such that

$$(X^{u,\theta})^n.h\ (x_1) \neq (X^{u,\theta})^n.h\ (x_2)$$

in other words be separated by a function of $\mathcal{O}$

■

Notice that observability is a global concept since it might be necessary to travel a considerable distance or for a long time to distinguish between points of $\mathbb{R}^n$. Therefore we introduce a local concept which is stronger than observability from [40]. Let U be a subset of $\mathbb{R}^n$ and $x_1, x_2 \in U$. We say x_1 is U -indistinguishable from x_2 (denoted $x_1\, I_U\, x_2$) if for every control u, whose trajectories $X_t^{\theta,u}(x_1)$ and $X_t^{\theta,u}(x_2)$ both lie in U, fails to distinguish between x_1 and x_2. U-indistinguishability is not, in general, an equivalence relation on U because it fails to be transitive.

However, we can still define (7.1) to be locally observable at x_0 if for every open neighborhood U of x_0, $\mathcal{I}_U(x_0) = \{x_0\}$ and locally observable if it is so on $\mathbb{R}^n$.

On the other hand one can weaken the concept of observability; in practice it may suffice to be able to distinguish x_0 from its neighbors. Therefore we define (7.1) to be weakly observable at x_0 if there exists a neighborhood U of x_0 such that $\mathcal{I}(x_0) \cap U = \{x_0\}$ and (7.1) is weakly observable if it is so on $\mathbb{R}^n$

Notice once again that it may be necessary to travel considerably far from U to distinguish points of U, so we make a last definition, (7.1) is locally weakly observable at x_0 if there exists an open neighborhood U of x_0 such that for every open neighborhood V of x_0 contained in U, $\mathcal{I}_V(x_0) = \{xo\}$ and is locally weakly observable if it is so in $\mathbb{R}^n$. Intuitively, (7.1) is locally weakly observable if one can instantaneously distinguish each point from its neighbors.

We have the following relations between these concepts [40] :

$$\begin{array}{ccc} \text{Locally Observable} & \Longrightarrow & \text{Observable} \\ \Downarrow & & \Downarrow \\ \text{Locally Weakly Observable} & \Longrightarrow & \text{Weakly Observable} \end{array}$$

7.2.2 Identifiability

Definition 7.2.4
The system (7.1) is said to be locally identifiable at θ if there exists a neighborhood N_θ of $\theta \in \mathbb{R}^p$ such that for any $\theta_1 \neq \theta_2$ in N_θ, there exists an admissible input u such that for a nonnegative time t we have $y(t, \theta_1, u) \neq y(t, \theta_2, u)$.
If $N_\theta = \mathbb{R}^p$ the system is said globally identifiable.

There is also a notion of identifiability almost everywhere. To introduce To introduce such a concept, we need a topology for the input function space. For any $T > 0$ and a positive integer n, the space $\mathcal{C}^n[0, T]$ is the space of all functions on $[0, T]$ which have continuous derivatives up to the order n. A topology of the space is the one associated with following well-defined norm, for $f \in \mathcal{C}^n[0, T]$, we define $\|f\| = \sum_{i=1}^{n} \max_{t\in[0,T]} \|f^{(i)}(t)\|$.

Definition 7.2.5
The system (7.1) is said to be structurally identifiable if there exists a time $T > 0$, a positive integer n, and open and dense subset of $\mathbb{R}^n$, $\mathbb{R}^p$ and $\mathcal{C}^n[0, T]$ such that the system is identifiable at θ for every x_0, θ and u in these open dense sets.

This characterizes the one-to-one property (almost everywhere) of the map from the parameters to the system output.

In these definitions, the initial state is unknown. However, sometimes an initial condition is known (or partially known). Then we can define x_0-identifiability easily, the output is then limited from a common initial state x_0.

7.2.3 Observability, identifiability and augmented system

The augmented system is simply the original system in an enlarged state-space with parameters treated as constant states :

$$\begin{cases} \dot{x} = & X(x, \theta, u) = X^{u,\theta}(x) \\ \dot{\theta} = & 0 \\ y = & h(x) \end{cases} \tag{7.2}$$

The system (7.1) is said observable and identifiable iff the enlarged system (7.2) is observable.
The algebraic identifiability is about construction of parameters from algebraic equations of the system input and output. This concept was f defined in [28] in the differential algebraic framework. We adapt the definition into the following one.

Definition 7.2.6
A system is said to be algebraically identifiable if it is possible to construct the parameters from solving algebraic equations depending only on the information of the input and output.

We will give a more precise definition by using the formalism of differential algebra. Differential algebra can be seen as a generalization to differential equations of the concepts of commutative algebra and algebraic geometry. This theory, founded by Ritt, is an appropriate framework for the definition of algebraic observability introduced by Diop and Fliess [28].
A differential ring R is a commutative ring, with an unity $1 \neq 0$ equipped with one derivation $R \to R$ $a \to \dot{a}$, such that $\overbrace{(a+b)}^{\cdot} = \dot{a} + \dot{b}$ and $\overbrace{(a.b)}^{\cdot} = \dot{a}.b + a.\dot{b}$. A constant of R is an element c such that $\dot{c} = 0$
Let K and L two differential fields such that $K \subset L$. If each element of L satisfies an algebraic differential equation with coefficients in K, then L is said to be differential algebraic extension of K. The differential field generated by K and a subset S of L is denoted by $K\langle S\rangle$.
Let $\mathbf{k}$ a differential field. Denote by$\mathbf{k}\langle u\rangle$ the differential field generated by $\mathbf{k}$ and a finite set $u = (u_1, \ldots, u_m)$ of differential quantities. The set u plays the role of control variables or input, which may be assumed to be independent. This means that u is differentially $\mathbf{k}$- algebraically independent.

A dynamic is is a finitely generated differential extension $\mathcal{D}/\mathbf{k}\langle u\rangle$

This means that any element of $\mathcal{D}$, satisfies an algebraic differential equation with coefficients which are rational functions over $\mathbf{k}$ in the components of u and a finite number of their derivatives. As output variables can be viewed as sensors on the dynamics, we formally define an output as a finite set $y = (y_1, \ldots, y_p) \in \mathcal{D}$

Definition 7.2.7
A dynamic $\mathcal{D}/\mathbf{k}\langle u\rangle$, with output y is said to be algebraically observable iff the differentail extension $\mathcal{D}/\mathbf{k}\langle u, y\rangle$ is algebraic

The intuitive meaning is the following: x the state variable can be expressed as an algebraic function of the components of $\{u, y\}$ and a finite number of their derivatives.

Definition 7.2.8
Parameters π are algebraically identifiable if they are observable with respect to $\{u, y\}$, i.e., if they are algebraic over $\mathbf{k}\langle u, y\rangle$. The parameters are rationally identifiable if they belongs to $\mathbf{k}\langle u, y\rangle$.

The intuitive meaning is clear: In the algebraic case, several values of the parameters are possible as algebraic equations must be solved, whereas in the rational case one is always insured of a single value. This rational identifiability is equivalent to the global identifiability of Glad and Ljung [65].
We will now give Theorems for algebraic identifiability [100].

Theorem 7.2.2 (Xia and Moog, 2003)
Let $\mathcal{K}$ the differential field consisting of meromorphix function of x, θ, u and finite derivatives of u and define $E = span_{\mathcal{K}}\{d\mathcal{K}\}$. A vector is in E if it is a finite linear combination of one forms dx, $d\theta$, du, $\ldots$, du^k, $\ldots$, with coefficients in $\mathcal{K}$.
Denote $\mathcal{Y} = \bigcup_{k=0}^{\infty} span\{dy, d\dot{y}, \cdots, dy^k\}$, $\mathcal{U} = \bigcup_{k=0}^{\infty} span\{du, d\dot{u}, \cdots, du^k\}$ and $\mathcal{X} = span\{dx\}$

The system is algebraically identifiable iff

$$\Theta \subset \mathcal{Y} + \mathcal{U}$$

Recall for a function h we define $\dot{h} = X^{u,\theta}.h$. Similarly for a one form ω in E if

$$\omega = \kappa_x\, dx + \kappa_\theta\, d\theta + \sum \eta_i\, du^i \in E$$

we define (Leibnitz rule)

$$\dot{\omega} = \dot{\kappa}_x\, dx + \dot{\kappa}_\theta\, d\theta + \sum \dot{\eta}_i\, du^i + \kappa_x\, X^{u,\theta}(x) + \sum \eta_i\, du^{i+1}$$

7.3 Examples

7.3.1 Identification for an intra-host model of Malaria

We consider the intra-host model of Malaria we have already encountered in these lectures

$$\begin{cases} \dot{x} &= \Lambda - \mu_x\, x - \beta\, x\, m, \\ \dot{y} &= \beta\, x\, m - \mu_y\, y, \\ \dot{m} &= r\, \mu_y\, y - \mu_m\, m - \beta\, x\, m. \end{cases} \tag{7.3}$$

Observability and Identifiability

In case of malaria, only the concentration of infected erythrocytes are measured [33, 35, 34]. For this system the parameters Λ, μ_x, μ_y, μ_m and r are known or at least widely accepted. However the infection rate is unknown. Then the problem, when y is measured, is to reconstruct the states x, m and identify the parameter β.

Proposition 7.3.1
System (7.3) , with observation $h(x, y, m) = y$ and parameters Λ, μ_x, μ_y, μ_m and r known, is observable and identifiable.

We assume, in relation to the plausible biological parameters, that we have the inequalities $\mu_x < \mu_y < \mu_m$ and $r > 1$.

Proof
We define $h_1 = y$. Then

$$\dot{y} = \beta\, x\, m - \mu_y\, y = \beta\, x\, m - \mu_y\, h_1.$$

We define the known function h_2 by $h_2 = \beta\, x\, m = \dot{h}_1 + \mu_y\, h_1$.
We now have

$$\dot{h}_2 = (\Lambda - h_2)\,\beta\, m + (r\,\mu_y\, h_1 - h_2)\,\beta\, x - (\mu_x + \mu_m)\, h_2$$

Again, setting $h_3 = (\Lambda - h_2)\,\beta\, m + (r\,\mu_y\, h_1 - h_2)\,\beta\, x$, we obtain

$$\dot{h}_3 = -\left[\mu_m\,(\Lambda - h_2) + \dot{h}_2\right]\,\beta\, m - \left[\mu_x\,(r\,\mu_y\, h_1 - h_2) - (r\,\mu_y\,\dot{h}_1 - \dot{h}_2)\right]\beta\, x \\ + 2\,\left[(\Lambda - h_2)\,(r\,\mu_y\, h_1 - h_2)\right]\,\beta$$

We define two known functions $A = \Lambda - h_2$ and $B = r\,\mu_y\; h_1 - h_2$.
Consider the two equations

$$\begin{cases} A\,\beta\, m + B\,\beta\, x & = -\dot{A} + (\mu_x + \mu_m)\, h_2 \\ (\mu_m\, A - \dot{A})\,\beta\, m + (\mu_x\, B - \dot{B})\,\beta\, x & = -\dot{h}_3 + 2\, A\, B\,\beta \end{cases} \tag{7.4}$$

These equations can be considered as a linear system in $\beta\, m$, $\beta\, x$ with constant terms depending on β. Let Δ the determinant of this system

$$\Delta = -A\,\dot{B} + \dot{A}\,B + (\mu_x - \mu_m)\, A\, B$$

The function Δ is analytic. Then either Δ is zero or its zeroes are isolated. We claim that Δ cannot be zero on a trajectory of system (7.3).

Moreover $A = \Lambda - h_2$ cannot be zero an a trajectory. If it is the case then $\dot{x} = -\mu_x\, x$ and x goes to zero, which is impossible for system (7.3) : x converges either to the first component of the DFE or the endemic equilibrium, which is, in any case, non zero. In the same spirit $B = r\,\mu_y\; h_1 - h_2$ cannot be zero on a trajectory, otherwise $\dot{y} = (r-1)\, y$ which is a contradiction with the boundedness of the trajectories of (7.3).

If $\Delta = 0$, dividing by $A\, B$ we have

$$-\frac{\dot{B}}{B} + \frac{\dot{A}}{A} = (\mu_m - \mu_x).$$

Integrating this relation gives

$$\ln \frac{B_0}{A_0}\,\frac{A}{B} = (\mu_m - \mu_x)\, t.$$

Where A_0 and B_0 are the initial value of A and B. Then

$$\frac{B_0}{A_0}\,\frac{A}{B} = e^{(\mu_m - \mu_x)\, t}$$

Since $\mu_m > \mu_x$ the right hand side of this relation converges infinity, and the left hand side converges to $\dfrac{B_0}{A_0}\,\dfrac{\mu_x\,\bar{x}}{\mu_m\,\bar{m}}$, where we denote by $(\bar{x}, \bar{y}\,\bar{m})$ either the DFE or the EE of system (7.3). In any case the left hand side does not converge to infinity, a contradiction.

Now with $\Delta \neq 0$ we obtain for $\beta\, x$ and $\beta\, m$

$$\begin{cases} \beta m &= M + N\,\beta \\ \\ \beta x &= P + Q\,\beta. \end{cases} \tag{7.5}$$

where M, N, P and Q are algebraic functions of the known functions h_1, h_2, $\dot{h}_1$,$\dot{h}_2$, $\dot{h}_3$.
From this relations we obtain, deriving one more time,

$$\begin{aligned} \beta\dot{m} &= \dot{M} + \dot{N}\,\beta \\ \beta\dot{m} &= (r\,\mu_y\, h_1 - h_2)\,\beta - \beta\,\mu_m\, m \\ \beta\dot{m} &= (r\,\mu_y\, h_1 - h_2)\,\beta - \mu_m(M + N\,\beta) \end{aligned} \tag{7.6}$$

Hence

$$\dot{M} + \dot{N}\,\beta = (r\,\mu_y\, h_1 - h_2)\,\beta - \mu_m(M + N\,\beta)$$

$$\left(B - \dot{N} - \mu_m\, N\right)\, \beta = \dot{M} + \mu_m\, M.$$

Since $M = \left[-A + (\mu_x + \mu_m)\, h_2\right](\mu_m\, A - \dot{A}) + A\, \dot{h}_3$, a similar argument, as the previous one, shows that $\dot{M} + \mu_m\, M$ cannot be zero on a trajectory. Finally we obtain an expression for β as a rational expression of the derivative of the measured output y. Note that we derive 4 times, corresponding to 3 states and one unknown parameter.

We prove the identifiability and observability of our system.

■

7.3.2 Identification for the Macdonald's model of schistosomiasis transmission

In this section we consider the very well known model of MacDonald for the transmission of Schistosomiasis [72, 11]
This model can be written as

$$\begin{cases} \dot{w} &= \alpha\, y - \gamma\, w \\ \\ \dot{y} &= \beta\, (1 - y)\, w - \mu\, y \end{cases} \tag{7.7}$$

The variable w is the average burden in the definitive host (e.g; humans) and the variable y is the prevalence of infection in snails. We assume that the average burden w is measured and that the transmission parameters α and β are unknown. γ the per capita death rate of parasites and μ the per capita death rate of infected snails are known.

Observability and identifiability

Proposition 7.3.2
System (7.7), with w measured, γ and μ known, is observable and identifiable, excepted at the equilibria $(0, 0)$ and $\left(\dfrac{\alpha}{\gamma}\left(1 - \dfrac{1}{\mathcal{R}_0}\right), 1 - \dfrac{1}{\mathcal{R}_0}\right)$

with $\mathcal{R}_0 = \dfrac{\alpha\, \beta}{\gamma\, \mu} > 1$

Proof
Observation is $h(w, y) = w$ then

$$\dot{h} = \alpha\, y - \gamma\, w \tag{7.8}$$
$$= \alpha\, y - \gamma\, h, \tag{7.9}$$

therefore

$$\alpha\, y = \dot{h} + \gamma\, h \tag{7.10}$$

$$\alpha\, \dot{y} = \alpha\, \beta\, (1 - y)\, h - \alpha\, \mu\, y \tag{7.11}$$
$$= \alpha\, \beta\, (1 - y)\, h - \mu\, (\dot{h} + \gamma\, h) \tag{7.12}$$
$$= \ddot{h} + \gamma\, \dot{h} \tag{7.13}$$

Function h is never identically zero (otherwise we are at the disease free equilibrium)

$$\alpha\, \beta\, (1 - y) = \frac{\ddot{h} + \gamma\, \dot{h} + \mu\, (\dot{h} + \gamma\, h)}{h} = g \tag{7.14}$$

Function g is known and is rationally expressed with the derivatives of h. We get

$$-\alpha\, \beta\, \dot{y} = \dot{g} \tag{7.15}$$
$$-\beta\, (\ddot{h} + \gamma\, \dot{h}) = \dot{g} \tag{7.16}$$

Parameter β has an rational expression in the derivatives of h, up to order 3.
If $\ddot{h} + \gamma\, \dot{h} = 0$ on a trajectory, we have $\dot{y} = 0$, hence y is constant, which gives w constant. We are on an equilibria. Since $h \neq 0$ we are at the endemic equilibrium.

$$w^* = \frac{\alpha}{\gamma}\, y^* \text{ and } y^* = 1 - \frac{\mu\, \gamma}{\alpha\, \mu} = 1 - \frac{1}{\mathcal{R}_0}$$

In this case the system is non identifiable (we only know the product $\alpha\, \beta$).
Otherwise

$$\beta = -\frac{\dot{g}}{\ddot{h} + \gamma\, \dot{h}} \tag{7.17}$$

Using (7.14) we obtain

$$\alpha\, \beta\, (1 - y) = \alpha\, \beta - \beta\, (\dot{h} + \gamma\, h) = g$$

which gives

$$\alpha = \frac{g + \beta\,(\dot{h} + \gamma\,h)}{\beta} \tag{7.18}$$

finally

$$y = \frac{\dot{h} + \gamma\,h}{\alpha} \tag{7.19}$$

System (7.7) is rationally observable and identifiable.

■

7.3.3 Identification for Ross' s model of malaria transmission

We consider the classical Ross model of malaria transmission

$$\begin{cases} \dot{x} = & \alpha\,y\,(1-x) - \gamma\,x \\ \\ \dot{y} = & \beta\,(1-y)\,x - \mu\,y \end{cases} \tag{7.20}$$

The variable x represents the prevalence of malaria in humans, variable y is the prevalence of infection in mosquitoes. Parameter γ is the recovery rate of infected humans and μ is the death rate of mosquitoes. Parameters α and β are composite parameter. Classically $\alpha = m\,a\,b_1$ where m is the vectorial density, i.e., mean number of mosquitoes by human, a is the biting rate and b_1 is the probability that a bite by an infectious mosquito infects a susceptible human. Similarly $\beta = a\,b_2$ with b_2 is the probability that a bite by a susceptible mosquito of an infected human gives an infected mosquito.
This is Ross model of 1911 [78, 4]. If we set $\alpha = m\,a\,b_1\,e^{-\mu\,\tau}$ with τ the mosquito incubation period, we obtain a version of Ross-Macdonald model.
Parameters γ and μ are known, at least locally, and α and β has to be identified. Again we assume that x is measured.

Observability and Identifiability

Proposition 7.3.3
Ross model (7.20) with x measured, parameters μ and γ known is observable and identifiable, excepted at the equilibria which are the disease free equilibrium $(0, 0)$ and the endemic equilibrium (if $\mathcal{R}_0 > 1$)

$$\left(\frac{\mathcal{R}_0 - 1}{\mathcal{R}_0 + \frac{\beta}{\mu}}, \frac{\mathcal{R}_0 - 1}{\mathcal{R}_0 + \frac{\alpha}{\gamma}}\right)$$

with $\mathcal{R}_0 = \dfrac{\alpha\,\beta}{\gamma\,\mu}$.

Proof
We set the observation $h_1(x, y) = x$. The function h_1 is never identically equal to 1, since if $y = 1$ then $\dot{y} = -\mu$. We define

$$h_2 = \alpha\, y = \frac{\dot{h}_1 + \gamma\, h_1}{1 - h_1}$$

Deriving h_2 gives

$$\dot{h}_2 = \alpha\,\dot{y} = \alpha\,\beta(1 - y)\,h_1 - \mu\,h_2 = \alpha\,\beta\,h_1 - \beta\,h_2\,h_1 - \mu\,h_2.$$

The function h_2 cannot be identically zero unless at the origin.
Then set

$$h_3 = \frac{\dot{h}_2 + \mu\,h_2}{h_1} = \alpha\,\beta - \beta\,h_2,$$

then

$$\dot{h}_3 = -\beta\,\dot{h}_2.$$

Excepted at one of the equilibria, $\dot{h}_2$ cannot be identically zero, therefore

$$\beta = -\frac{\dot{h}_3}{\dot{h}_2}.$$

Finally

$$\alpha = \frac{h_3 + \beta\,h_2}{\beta} \text{ and } y = \frac{h_2}{\alpha}$$

■

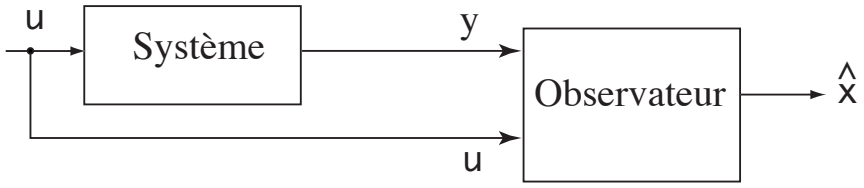
7.4 Identifiability and observers

7.4.1 Definition

When a system is observable initial points are distinguishable. Is it possible to reconstruct the state of the system from the measurement ? Observability deems to be a natural requirement.

We consider an inout-output system

$$\begin{cases} \dot{x} = & X(x,u) = X^u(x) \\ y = & h(x) \end{cases}$$



From control theory, a system which gives an estimation of the state, is called an observer. To be more precise an observer is a dynamical system, whose input are the known quantities fo the control system, i.e., u and y :

$$\dot{\hat{x}} = g(\hat{x}, u, u),$$

with the property that

$$\lim_{t \to +\infty} \|\hat{x}(t) - x(t)\| = 0$$

The observer is called an asymptotic observer if the error satisfies for constants $K > 0$ $a > 0$ and for ant $t > 0$

$$\|\hat{x}(t) - x(t)\| \leq K \, \|\hat{x}(0) - x(0)\| \, e^{-a\,t}$$

7.4.2 An example : within-host model of Malaria

We have already seen this type of model. The observation is the concentration of infected red blood cells. Actually what is observed are the young parasites in peripheral blood. This the phenomenon of sequestration : At the half-way point of parasite development, the infected erythrocyte leaves the circulating blood and binds to endothelium in the microvasculature.

Then we will distinguish two types of infected RBC

$$\begin{cases} \dot{x} = \Lambda - \mu_x\, x - \beta x\, m \\ \\ \dot{y}_1 = \beta x\, m - \alpha_1\, y_1 \\ \\ \dot{y}_2 = \gamma_1\, y_1 - \alpha_2\, y_2 \\ \\ \dot{m} = r\, \gamma_2\, y_2 - \mu_m\, m - \beta\, x\, m \end{cases} \tag{7.21}$$

We observe y_1 at discrete times. y_2 is the population of sequestered infected erythrocytes. Can we estimate the system state (x, y_1, y_2, m) ? The answer is yes : [16].
We rewrite the system in the following form

$$\begin{cases} \dot{z} = A\, z + \Lambda\, e_1 + d(t)\, E \\ \\ \mathcal{Y} = C\, z \end{cases} \tag{7.22}$$

where $d(t)\, E = \beta\, x\, m$ is considered as an unknown input

$$A = \begin{bmatrix} -\mu_x & 0 & 0 & 0 \\ 0 & -\alpha_1 & 0 & 0 \\ 0 & \gamma_1 & -\alpha_2 & 0 \\ 0 & 0 & r\,\gamma_2 & -\mu_m \end{bmatrix} \quad E = \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \end{bmatrix} \quad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$$

Let $\bar{A}$ be

$$\bar{A} = (I_4 - E\, C)\, A$$

Then the pair $(\bar{A}, C)$ is detectable, i.e.,

$$rank[C^T, \bar{A}^T\, C^T, (\bar{A}^T)^2\, C^T, (\bar{A}^T)^3\, C^T] = 4$$

This implies that there exists a 4×1 matrix L such that $\bar{A} - L\, C$ is Hurwitz. This can be verified easily Let $L = [l - 1, l_2, l_3, l_4]$ then

$$\bar{A} - L\, C = \begin{bmatrix} -\mu_x & -\alpha_1 - l_1 & 0 & 0 \\ 0 & -l_2 & 0 & 0 \\ 0 & \gamma_1 - l_3 & -\alpha_2 & 0 \\ 0 & -\alpha_1 - l_4 & \gamma_2 & -\mu_m \end{bmatrix}$$

Eigenvalues of this matrix are $-\mu_x$, $-l_2$, $-\alpha_2$ a,d $-\mu_m$. It is sufficient to choose $l_2 > 0$ to have $\bar{A} - L\, C$ is Hurwitz.
With this condition we claim that

$$\begin{cases} \dot{w} = (\bar{A} - L\, C)\, w + \left[L + (\bar{A} - L\, C)\, E\right]\, \mathcal{Y} + \Lambda\, e_1 \\ \\ \hat{z} = w(t) + E\, \mathcal{Y} \end{cases},$$

To prove this claim we set the error $e = \hat{z} - z$. Some remarks are in order

$$E\, C\, E = E \text{ and } C\, e_1 = 0$$

$$\begin{aligned} \dot{e} &= \dot{\hat{z}} - \dot{z} \\ &= \dot{w} + E\, \dot{\mathcal{Y}} - \dot{z} \\ &= \dot{w} + E\, \dot{\mathcal{Y}} - A\, z - \Lambda\, e_1 - d(t)\, E \\ &= \dot{w} + E\, \dot{\mathcal{Y}} - (I - L\, C)\, A\, z - E\, C\, \left[A\, z + \Lambda\, e_1 + d(t)\, E\right] - \Lambda\, e_1 \\ &= \dot{w} + E\, \dot{\mathcal{Y}} - \bar{A}\, z - E\, \dot{\mathcal{Y}} - \Lambda\, e_1 \\ &= \dot{w} - \bar{A}\, z - \Lambda\, e_1 \\ &= (\bar{A} - L\, C)\, w + \left[L + (\bar{A} - L\, C)\, E\right]\, \mathcal{Y} + \Lambda\, e_1 - \bar{A}\, z - \Lambda\, e_1 \\ &= (\bar{A} - L\, C)\, w + \left[L + (\bar{A} - L\, C)\, E\right]\, \mathcal{Y} - \bar{A}\, z \\ &= (\bar{A} - L\, C)\, \hat{z} + L\, \mathcal{Y} - \bar{A}\, z \\ &= (\bar{A} - L\, C)\, \hat{z} + L\, C\, z - \bar{A}\, z \\ &= (\bar{A} - L\, C)\, (\hat{z} - z) \\ &= (\bar{A} - L\, C)\, e \end{aligned}$$

Since $\bar{A} - L\, C$ is Hurwitz this proves the claim. Our observer is asymptotic. The speed of convergence is given by the stability modulus, namely $-\mu_x$ if l_2 is chosen large enough.

This observer is known as an observer with unknown input. The state of the system is estimated without knowledge of the parameter β. Asa by-product we can now estimates this parameter. For this

We have

$$y_1 = \beta\, x\, m - (\gamma_1 + \mu_1)\, y_1$$

Constant variation formula for linear system gives

$$y_1(t) = \exp\left(-(\gamma_1+\mu_1)(t-t_0)\right)\, y_1(t_0) + \beta \int_{t_0}^{t} x(s)\, m(s)\, \exp\left(-(\gamma_1+\mu_1)(s-t)\right)\, ds$$

We replace the unknown values by their estimations

$$\hat{y}_1(t) = \exp\left(-(\gamma_1+\mu_1)(t-t_0)\right)\, \hat{y}_1(t_0) + \hat{\beta} \int_{t_0}^{t} \hat{x}(s)\, \hat{m}(s)\, \exp\left(-(\gamma_1+\mu_1)(s-t)\right)\, ds$$

therefore

$$\exp\left((\gamma_1+\mu_1)t\right)\, \hat{y}_1(t) - \exp\left((\gamma_1+\mu_1)\, t_0\right)\, \hat{y}_1(t_0) = \hat{\beta} \int_{t_0}^{t} \hat{x}(s)\, \hat{m}(s)\, \exp\left(-(\gamma_1+\mu_1)\, s\right)\, ds$$

Discretizing $[t_{\text{init}}, t_f]$ in $[t_i, t_{i+1}]$ applying preceding relation and choosing t_{init} such that the observer gives a good estimation (practically that means waiting a little, giving time for the observer to converge

$$\exp\left((\gamma_1+\mu_1)t_{i+1}\right)\, \hat{y}_1(t_{i+1}) - \exp\left((\gamma_1+\mu_1)\, t_i\right)\, \hat{y}_1(t_i) = \hat{\beta} \int_{t_i}^{t_{i+1}} \hat{x}(s)\, \hat{m}(s)\, \exp\left(-(\gamma_1+\mu_1)\, s\right)\, ds$$

For each i we set

$$U_i = \exp\left((\gamma_1+\mu_1)t_{i+1}\right)\, \hat{y}_1(t_{i+1}) - \exp\left((\gamma_1+\mu_1)\, t_i\right)\, \hat{y}_1(t_i)$$

and

$$V_i = \int_{t_i}^{t_{i+1}} \hat{x}(s)\, \hat{m}(s)\, \exp\left(-(\gamma_1+\mu_1)\, s\right)\, ds$$

this gives

$$U = \hat{\beta}\, V$$

An estimation of β is obtained in solving by least square method this linear system.

Malariatherapy

During the pre-penicillin era, there was no efficient treatment for syphilis. During the natural evolution of the disease, patients sometimes developed neurosyphilis 10 to 25 years after the initial infection. The curative effect of fevers has been reported since Hippocrates' time. Wagner-Jauregg, the father of malaria therapy, described in detail experiments with induced fever in patients.
In 1917 the inoculation of malaria parasites, which proved to be very successful in the case of dementia paralytica (also called general paresis of the insane), caused by neurosyphilis, at that time a terminal disease. It had been observed that some who develop high fevers could be cured of syphilis. Thus, from 1917 to the mid 1940s, malaria induced by the least aggressive parasite, *Plasmodium vivax*, was used as treatment for tertiary syphilis because it produced prolonged and high fevers (a form of pyrotherapy). This was considered an acceptable risk because the malaria could later be treated with quinine, which was available at that time. This discovery earned to Wagner-Jauregg the Nobel Prize in Medicine in 1927. The technique was known as malariatherapy; however, it was dangerous, killing about 15% of patients, so it is no longer in use.
Data were collected by the US Public Health Service between 1940 and 1963, when malaria therapy was a recommended treatment for neurosyphilis. Infections with different strains of *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae* were induced in patients for the treatment of neurosyphilis.
Afroamericans were however found to be refractory, and so they were treated with different strains of *P. falciparum* under close medical supervision. They were inoculated either with sporozoites (generally through mosquito bite) or with infected blood. Inoculations were preceded by variable sequences of blood and mosquito passages of the strain. Microscopic examination of the blood was performed on an almost daily basis.

Is the malaria therapy model identifiable ?

When the initial condition is partially known the system can be observable and identifiable [93, 100].
If we consider the case of malariatherapy the initial condition is partially known. Sporozoites are introduced and measured was performed daily.

Hence $(x_0 = \dfrac{\Lambda}{\mu}, y_2(0) = 0, m(0) = 0$ are known. Only the value of $y_1(0)$ is unknown.

Now $y_1(t)$ is known, hence $\dot{y}_1 + \alpha_1\, y_1 = \beta\, x\, m$ is also known. Consequently

$$x(t) = e^{-\mu_x\, t} \int_0^t (\Lambda - \beta\, x(s)\, m(s))\, ds + e^{-\mu_x\, t}\, x_0,$$

is also known.Then we know $\beta\, m$. Similarly

$$y_2(t) = e^{-\alpha_2\, t}\, \gamma_1 \int_0^t y_1(s)\, ds + e^{-\alpha_2\, t}\, y_2(0) = e^{-\alpha_2\, t}\, \gamma_1 \int_0^t y_1(s)\, ds$$

is known. Finally

$$\begin{aligned} m(t) &= e^{-\mu_m\, t} \int_0^t \left[r_2\, \gamma_2\, y_2(s) - \beta\, x(s)\, m(s)\right]\, ds + e^{-\mu_m\, t}\, m(0) \\ &= e^{-\mu_m\, t} \int_0^t \left[r_2\, \gamma_2\, y_2(s) - \beta\, x(s)\, m(s)\right]\, ds, \end{aligned}$$

is known. Since we can know x, y_1, y_2, m and $\beta\, x\, m$, $beta$is known.
This proves that the system (7.21) is observable and identifiable.

7.4.3 Numerical observers

If we prove that the system under consideration is observable and identifiable how can we reconstruct the states and the unknown parameters. Dynamical observers can be designed, but it is a difficult task. recently some advances have been obtained [16, 24, 51, 100]. However we will propose something else.
We will consider observability, since we have seen that the identifiability can be, with the augmented system, analyzed as an observability problem.

$$\left\{ \begin{array}{ll} \dot{x} & = f(x,u) \\ y & = h(x) \end{array} \right.$$

Measures are discrete, with a time interval of Δt :

$$(y_0, y_1, \ldots, y_k),$$

with corresponding time

$$(0, \Delta t, \cdots, k\, \Delta t)$$

The principle is very simple : for a initial condition (this is a guess) we compute the ouptput

$$(h(x_0), h(\, x(\Delta t, x_0, u)\,), \cdots, h(\, x(k\, \Delta t, x_0, u)\,).$$

Then is evaluated the difference between what is predicted and what is measured

$$\Phi(x_0) = \min_{x_0} \sum_{i=0}^{k} \|y_i - y(i\,\Delta t, x_0, u)\|_2^2.$$

This a function of x_0. Using an algorithm of minimization we compute

$$\hat{x}_0 = \arg\min_{x_0} \Phi(x_0)$$

By a numerical integration we obtain

$$\hat{x}(t) = x(t, \hat{x}_0, u)$$

As the number of measurements grows, the size of the optimization increases. To bound the size of the optimization, the least squares objective can be modified to employ a fixed-size moving window in which the number of measurements that we base our estimate on (and hence the size of the optimization) remains constant. Practically that means that, if our horizon is of length N, we will estimate x_0 using the N first measures. The we will have an estimation $\hat{x}(N\,\Delta t)$. This estimation will be the guess for the next time window $[(N+1)\,\Delta t, \cdots, 2\,N\,\Delta t)]$ giving a new $\hat{x}(N\,\Delta t)$ and an estimation for $x(2\,N\,\Delta t)$ ....
The use of a finite horizon and moving windows is motivated by two reasons. The first one is the limited amount of memory, which limits the size of the optimization problem. The second one is that generally measures are corrupted by noise. Hence taking many measures will add many noise to the data used in computation.
What is the size of the window ? This is an engineering problem, but we have some hints. E. Sontag has proved that For differential equations with r parameters $2\,r+1$ experiments are enough for identification. Aeyels has shown that in n-dimensions $2\,n+1$ samples are necessary to observe observability [88, 1, 2]. Then for the augmented system with n states and p parameters the size of the window should not be less than $2\,(n+p)+1$.

A program

Here a routine, written in Scilab (like MATLAB but free)

```
1 function Xob=numobs3(dyn,output,Tm,N,x0_est, x0_reel)
2 // Numerical observer
3 // comparion with real measures
4 //
5
6 // dyn  dynamical of system
7 // output :observation
```

```
//(vectorialized)
// The function  dyn must be defined
// as the output function

// Tm  vector of time measures
// N number of points in window
// x0_guess  initial guess
// x0_reel   real initial condition
// This program compute real trajectories and
// estimated one

//  k number of windows
// windows are joining at their extremities
k=floor((length(Tm)-1)/(N-1))

 //  initialzsation
 //real values
 Xreel=ode(x0_reel,Tm(1),Tm,dyn)
 //curves
 XXreel=ode(x0_reel,Tm(1),linspace(Tm(1),Tm($),1000),dyn)
 // observer on frist window
 Xest=ode(x0_est,Tm(1),Tm(1:N),dyn)
 Mesures=output(Xreel)
 // Measures with noise
 Mesuresb=Mesures+grand(1,length(Mesures),'nor',0,1)
nb_etat=length(x0_est) // state dimension
x0_guess=x0_est
// intial conditions in memory
//in each window
X=[]
// curves points in memory
XX=[]
// curves times in memory
TT=[]
//
for i=1:k

  z=1+(i-1)*(N-1):1+i*(N-1) // index window  i

  T=Tm(z);
```

```
48    mesureF=Mesuresb(z)
49    x0_opt=optimize(dyn,output,mesureF,x0_guess)
50    Z=ode(x0_opt,T(1),T,dyn)
51    sol=ode(x0_opt,T(1),linspace(T(1),T($)),dyn);
52   TT=[TT,linspace(T(1),T($))]
53    XX=[XX,sol]
54    X=[X,Z(:,1:$-1)]
55    x0_guess=Z(:,$)
56
57  end
58
59  // If the windows has not taken all the points
60  // one last window
61
62  if pmodulo(length(Tm),N-1)~=1
63    T=Tm(1+k*(N-1):$)
64    mesureFi=Mesuresb(1+k*(N-1):$)
65    x0_opt=optimize(dyn,output,mesureFi,x0_guess)
66    Z=ode(x0_opt,T(1),T,dyn)
67    sol=ode(x0_opt,T(1),linspace(T(1),T($)),dyn);
68    TT=[TT,linspace(T(1),T($))]
69    XX=[XX,sol]
70    X=[X,Z(:,1:$)]
71  end
72
73
74  Xob=X;
75
76  for j=1:nb_etat
77    xset("window",j)
78    clf
79    plot(Tm',[Xreel(j,:);Xob(j,:)]','+')
80    plot(TT',XX(j,:)','b')
81    plot(linspace(Tm(1),Tm($),1000)',XXreel(j,:)','g:')
82  end
83
84  xset("window",nb_etat+1)
85  clf
86  plot(Tm',[Mesures;Mesuresb]','o')
87
```

```
endfunction

//////////////////////////////////////
function x0_opt=optimize(fun,output,mesureF,x0_guess,T)
m=length(mesureF)
x0_opt=lsqrsolve(x0_guess,crit,m)
endfunction

//////////////////////
function y=crit(x,m)
z=ode(x,T(1),T,fun);
y=output(z)
y=y-mesureF
y=y(:)
endfunction
```

Example for the Schistosome example

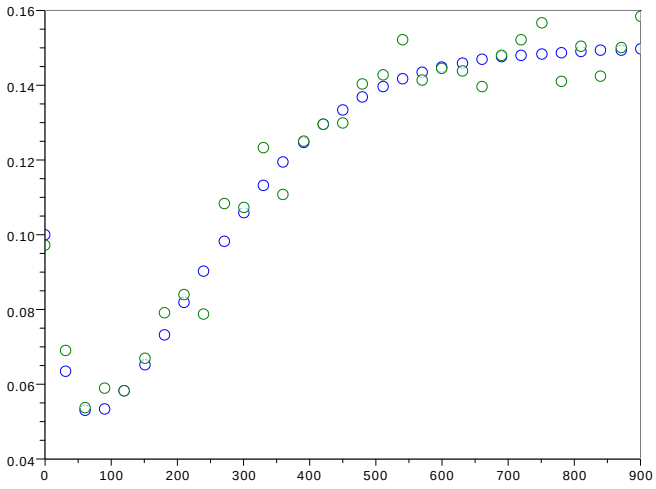


Figure 7.3: Measures w : blue real measures, green with noise

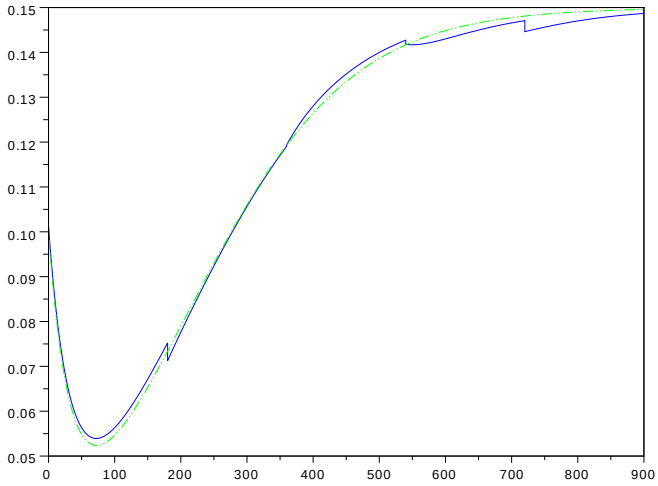


Figure 7.4: Estimation of w : noise is filtered

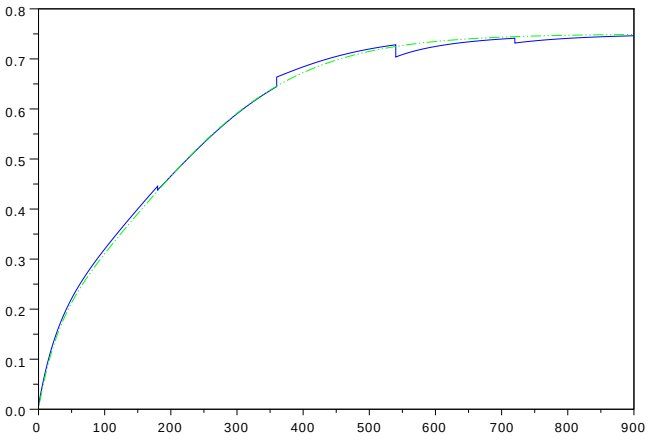


Figure 7.5: Estimation of y

Bibliography

[1] D. Aeyels, *Generic observability of differentiable systems*, SIAM J. Control Optim., 19 (1981), pp. 595–603.

[2] ——, *On the number of samples necessary to achieve observability*, Systems Control Lett., 1 (1981/82), pp. 92–94.

[3] R. M. Anderson and R. M. May, *Directly transmitted infections diseases: control by vaccination.*, Science, 215 (1982), pp. 1053–1060.

[4] ——, *Infectious Diseases of Humans. Dynamics and Control*, Oxford science publications, 1991.

[5] R. Anguelov, Y. Dumont, and J. Lubuma, *Mathematical modeling of sterile insect technology for control of anopheles mosquito*, Comput. Math. Appl., 64 (2012), pp. 374–389.

[6] J. Arino, J. Davis, D. Hartley, R. Jordan, J. Miller, and P. van den Driessche, *A multi-species epidemic model with spatial dynamics*, Math. Med. Biol., 22 (2005), pp. 129–142.

[7] J. Arino, R. Jordan, and P. van den Driessche, *Quarantine in a multi-species epidemics model with spatial dynamics*, Math. Biosci., (2006).

[8] J. Arino and P. van den Driessche, *Disease spread in metapopulations*, in Nonlinear dynamics and evolution equations, X.-O. Zhao and X. Zou, eds., vol. 48, Fields Instit. Commun., AMS, Providence, R.I., 2006, pp. 1–13.

[9] K. Arrow, *A "Dynamic" Proof of the Frobenius-Perron Theorem for Metzler Matrices*, in Probability, Statistics and Mathematics. Papers in honor of Samuel Karlin, Academic Press, 1989, pp. 17–25.

[10] N. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, Griffin,London, 1975.

[11] A. D. Barbour, *Macdonald's model and the transmission of bilharzia.*, Trans R Soc Trop Med Hyg, 72 (1978), pp. 6–15.

[12] Bellman, R. and Åström, K.J., *On structural identifiability*, Math. Biosci., 7 (1970), pp. 329–339.

[13] E. Beretta and Y. Takeuchi, *Global stability of Lotka-Volterra diffusion models with continuous time delay.*, SIAM J. Appl. Math., 48 (1988), pp. 627–651.

[14] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, vol. 9 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Revised reprint of the 1979 original.

[15] N. P. Bhatia and G. P. Szegö, *Dynamical systems: Stability theory and applications*, vol. 35 of Lecture Notes in Mathematics, Springer-Verlag, 1967.

[16] D. Bichara, N. Cozic, and A. Iggidr, *On the estimation of sequestered infected erythrocytes in plasmodium falciparum malaria patients*, Math. Biosci. Eng., 11 (2014), pp. 741–759.

[17] F. Brauer and C. Castillo-Chávez, *Mathematical models in population biology and epidemiology*, vol. 40 of Texts in Applied Mathematics, Springer-Verlag, New York, 2001.

[18] F. Brauer, J. Wu, and P. van den Driessche, eds., *Mathematical Epidemiology*, no. 1945 in Lectures Notes in Math., Springer-Verlag, 2008.

[19] G. Butler and P. Waltman, *Persistence in dynamical systems*, J. Differential Equations, 63 (1986), pp. 255–263.

[20] K. Cooke, P. van den Driessche, and X. Zou, *Interaction of maturation delay and nonlinear birth in population and epidemic models.*, J Math Biol, 39 (1999), pp. 332–352.

[21] K. L. Cooke and Z. Grossman, *Discrete delay, distributed delay and stability switches*, J. Math. Anal. Appl., 86 (1982), pp. 592–627.

[22] D. Cox and H. Miller, *The theory of stochastic processes*, Chapman and Hall, 1965.

[23] P. De Leenheer and H. L. Smith, *Virus dynamics: A global analysis.*, SIAM J. Appl. Math., 63 (2003), pp. 1313–1327.

[24] M. Diaby, A. Iggidr, and M. Sy, *Observer design for a schistosomiasis model*, Math. Biosci., 269 (2015), pp. 17–29.

[25] O. Diekmann and J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases*, Wiley Series in Mathematical and Computational Biology, John Wiley & Sons Ltd., Chichester, 2000. Model building, analysis and interpretation.

[26] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol., 28 (1990), pp. 365–382.

[27] O. Diekmann and M. Kretzschmar, *Patterns in the effects of infectious diseases on population growth.*, J Math Biol, 29 (1991), pp. 539–570.

[28] S. Diop and M. Fliess, *Nonlinear observability, identifiability, and persistent trajectories*, in proceedings 36^{th} IEEE-CDC, 1991, pp. 714–719.

[29] Y. Dumont and J. M. Tchuenche, *Mathematical studies on the sterile insect technique for the chikungunya disease and aedes albopictus.*, J Math Biol, (2011).

[30] A. Fall, A. Iggidr, G. Sallet, and J. J. Tewa, *Epidemiological models and Lyapunov functions*, Math. Model. Nat. Phenom., 2 (2007), pp. 55–73.

[31] D. Fargue, *Réductibilité des systèmes héréditaires à des systèmes dynamiques.*, C. R. Math. Acad. Sci. Ser B, 277 (1973), pp. 471–473.

[32] M. Fliess, *Nonlinear control theory and differential algebra*, in Modelling and adaptive control (Sopron, 1986), vol. 105 of Lect. Notes Control Inf. Sci., Springer, Berlin, 1988, pp. 134–145.

[33] M. B. Gravenor, A. L. Lloyd, P. G. Kremsner, M. A. Missinou, M. English, K. Marsh, and D. Kwiatkowski, *A model for estimating total parasite load in falciparum malaria patients.*, J Theor Biol, 217 (2002), pp. 137–48.

[34] M. B. Gravenor, A. R. McLean, and D. Kwiatkowski, *The regulation of malaria parasitaemia: parameter estimates for a population model.*, Parasitology, 110 (Pt 2) (1995), pp. 115–22.

[35] M. B. Gravenor, M. B. van Hensbroek, and D. Kwiatkowski, *Estimating sequestered parasite population dynamics in cerebral malaria.*, Proc Natl Acad Sci U S A, 95 (1998), pp. 7620–4.

[36] Z. Grossman, M. Feinberg, K. Kuznetsov, D. Dimitrov, and W. Paul, *Hiv infection : how effective is drug combination treatment*, Immunol. Today, 19 (1998), pp. 528–532.

[37] J. Hale, *Ordinary differential equations*, Krieger, 1980.

[38] J. A. Heesterbeek and J. A. Metz, *The saturating contact rate in marriage- and epidemic models.*, J Math Biol, 31 (1993), pp. 529–539.

[39] J. A. P. Heesterbeek, *A brief history of R_0 and a recipe for its calculation*, Acta Biotheorica, 50 (2002), pp. 189–204.

[40] R. Hermann and A. J. Krener, *Nonlinear controllability and observability.*, IEEE Trans. Autom. Control, 22 (1977), pp. 728–740.

[41] H. W. Hethcote, *The mathematics of infectious diseases*, SIAM Rev., 42 (2000), pp. 599–653 (electronic).

[42] H. W. Hethcote and H. R. Thieme, *Stability of the endemic equilibrium in epidemic models with subpopulations*, Math. Biosci., 75 (1985), pp. 205–227.

[43] M. Hirsch, *Systems of differential equations that are competitive or cooperative i: limit sets*, SIAM J. Math. Anal., 13 (1982), pp. 167–179.

[44] ——, *The dynamical system approach to differential equations*, Bull. AMS, 11 (1984), pp. 1–64.

[45] ——, *Systems of differential equations that are competitive or cooperative ii : convergence almost everywhere*, SIAM J. Math. Anal., 16 (1985), pp. 423–439.

[46] ——, *Systems of differential equations that are competitive or cooperative iii: Competing species*, Nonlinearity, 1 (1988), pp. 51–71.

[47] ——, *Systems of differential equations that are competitive or cooperative. iv: Structural stability in three-dimensional systems*, SIAM J. Appl. Math., 21 (1990), pp. 1125–1234.

[48] M. Hirsch and S. Smale, *Differential equations, dynamical systems, and linear algebra*, Pure and applied mathematics, Academic Press, 1974.

[49] M. W. Hirsch and H. Smith, *Monotone dynamical systems*, in Handbook of differential equations: ordinary differential equations. Vol. II, Elsevier B. V., Amsterdam, 2005, pp. 239–357.

[50] A. Iggidr, J. Mbang, and G. Sallet, *Stability analysis of within-host parasite models with delays*, Math. Biosci., 209 (2007).

[51] A. Iggidr and M. Souza, *State estimators for some epidemiological systems*, J. Math. Biol., to appear (2018).

[52] J. A. Jacquez, *Compartmental analysis in Biology and Medicine*, BioMedware, 1996.

[53] J. A. Jacquez and C. P. Simon, *Qualitative theory of compartmental systems*, SIAM Rev., 35 (1993), pp. 43–79.

[54] ——, *Qualitative theory of compartmental systems with lags*, Math. Biosci., 180 (2002), pp. 329–362.

[55] R. E. Kalman, *Mathematical description of linear dynamical systems*, J. SIAM Control Ser. A, 1 (1963), pp. 152–192 (1963).

[56] J. P. Keener, *The Perron-Frobenius theorem and the ranking of football teams*, SIAM Rev., 35 (1993), pp. 80–93.

[57] W. Kermack and A. McKendrick, *A contribution to the mathematical theory of epidemics*, Proc. R. Soc., A115 (1927), pp. 700–721.

[58] Krasnosel'skiĭ, M. A., *Positive solutions of operator equations*, Translated from the Russian by Richard E. Flaherty; edited by Leo F. Boron, P. Noordhoff Ltd. Groningen, 1964.

[59] Krasnosel′skiĭ, M. A., *The operator of translation along the trajectories of differential equations*, Translations of Mathematical Monographs, Vol. 19. Translated from the Russian by Scripta Technica, American Mathematical Society, Providence, R.I., 1968.

[60] A. Lajmanovich and J. Yorke, *A deterministic model for gonorrhea in a nonhomogeneous population.*, Math. Biosci., 28 (1976), pp. 221–236.

[61] J. P. LaSalle, *The stability of dynamical systems*, CBMS conferences, 25, SIAM, Philadelphia, Pa., 1976. With an appendix: "Limiting equations and stability of nonautonomous ordinary differential equations" by Z. Artstein, Regional Conference Series in Applied Mathematics.

[62] A. Laveran, , A. Bull.Acad. N. Med, 9 (1880), p. 1235.

[63] E. Lee and L. Markus, *Foundations of optimal control theory*, John Wiley & Sons Ltd., 1967.

[64] M. Li, *An introduction to Mathematical Modeling of Infectious diseases*, Springer, 2018.

[65] L. Ljung and T. Glad, *On global identifiability for arbitrary model parametrizations*, Automatica J. IFAC, 30 (1994), pp. 265–276.

[66] A. L. Lloyd, *The dependance of viral parameter estimates on the assumed viral life cycle : limitations of studies of viral load data*, Proc R Soc Lond B Biol Sci, (2001), pp. 847–854.

[67] A. L. Lloyd, *Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods.*, Proc R Soc Lond B Biol Sci, 268 (2001), pp. 985–93.

[68] ——, *Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics.*, Theor Popul Biol, 60 (2001), pp. 59–71.

[69] A. Lotka, *contribution to the analysis of malaria epidemiology*, Am. J. Trop. Med. Hyg., 3 (1923), pp. 1–121.

[70] Z. Y. Lu and Y. Takeuchi, *Global asymptotic behavior in single-species discrete diffusion systems*, J. Math. Biol., 32 (1993), pp. 67–77.

[71] D. G. Luenberger, *Introduction to dynamic systems. Theory, models, and applications.*, John Wiley & Sons Ltd., 1979.

[72] G. Macdonald, *The dynamics of helminth infections, with special reference to schistosomes.*, Trans R Soc Trop Med Hyg, 59 (1965), pp. 489–506.

[73] N. MacDonald, *Time lags in biological models*, no. 27 in Lecture Notes in Biomath., Springer-Verlag, 1978.

[74] J. E. Mittler, B. Sulzer, A. U. Neumann, and A. S. Perelson, *Influence of delayed viral production on viral dynamics in HIV-1 infected patients.*, Math. Biosci., 152 (1998), pp. 143–163.

[75] P. W. Nelson and A. S. Perelson, *Mathematical analysis of delay differential equation models of HIV-1 infection.*, Math. Biosci., 179 (2002), pp. 73–94.

[76] M. A. Nowak and R. M. May, *virus dynamics. Mathematical principles of immunology and virology*, Oxford University Press, 2000.

[77] R. Ross, *Report on the prevention of Malaria in Mauritius*, Waterloo and sons, London, 1908.

[78] R. Ross, *The prevention of malaria*, John Murray, 1911.

[79] R. Ross, *Some quantitative studies in epidemiology*, Nature, 87 (1911), pp. 466–467.

[80] L. Schwartz, *Methods of mathematical physics*, Addison-Wesley, Reading Mass., 1964.

[81] P. Seibert and R. Suarez, *Global stabilization of nonlinear cascade systems*, Syst. Control Lett., 14 (1990), pp. 347–352.

[82] J. F. Selgrade, *Asymptotic behavior of solutions to single loop positive feedback systems*, J. Differential Equations, 38 (1980), pp. 80–103.

[83] E. Seneta, *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York, 2006. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544].

[84] H. Smith and P. Waltman, *The theory of the chemostat. Dynamics of microbial competition.*, Cambridge Studies in Mathematical Biology, Cambridge University Press, 1995.

[85] H. L. Smith, *Cooperative systems of differential equations with concave nonlinearities*, Nonlinear Anal., 10 (1986), pp. 1037–1052.

[86] ——, *Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems.*, Mathematical Surveys and Monographs. 41. Providence, RI: American Mathematical Society (AMS). x, 174 p. , 1995.

[87] E. D. Sontag, *Mathematical control theory, deterministic finite dimensional systems*, no. 6 in Texts in Applied Mathematics, Springer-Verlag, 1990.

[88] ——, *For differential equations with r parameters, $2r + 1$ experiments are enough for identification*, J. Nonlinear Sci., 12 (2002), pp. 553–583.

[89] Y. Takeuchi, *Cooperative systems theory and global stability of diffusion models*, Acta Appl. Math., 14 (1989), pp. 49–57. Evolution and control in biological systems (Laxenburg, 1987).

[90] ——, *Global dynamical properties of Lotka-Volterra systems*, World Scientific Publishing Co. Inc., River Edge, NJ, 1996.

[91] H. R. Thieme, *Persistence under relaxed point-dissipativity (with applications to an endemic model)*, Siam J. Math. Anal., 24 (1993), pp. 407–435.

[92] H. R. Thieme, *Mathematics in population biology*, Princeton Series in Theoretical and Computational Biology, Princeton University Press, Princeton, NJ, 2003.

[93] E. T. Tunali and T. J. Tarn, *New results for identifiability of nonlinear systems*, IEEE Trans. Automat. Control, 32 (1987), pp. 146–154.

[94] P. van den Driessche and J. Watmough, *reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci., (2002), pp. 29–48.

[95] R. Varga, *Factorization and normalized iterative methods*, in Boundary problems in differential equations, R. Langer, ed., university of Wisconsin Press, 1960, pp. 121–142.

[96] ——, *matrix iterative analysis*, Prentice-Hall, 1962.

[97] M. Vidyasagar, *Decomposition techniques for large-scale systems with nonadditive interactions: Stability and stabilizability.*, IEEE Trans. Autom. Control, 25 (1980), pp. 773–779.

[98] P. Waltman, *Competition models in population biology*, vol. 45 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1983.

[99] W. Wang and X.-Q. Zhao, *An epidemic model in a patchy environment*, Math. Biosci., 190 (2004).

[100] X. Xia and C. H. Moog, *Identifiability of nonlinear systems with application to HIV/AIDS models*, IEEE Trans. Automat. Control, 48 (2003), pp. 330–336.