Controllability, Observability, and Stability of Mathematical Models

Abderrahman Iggidr

6.3.1.4 CONTROLLABILITY, OBSERVABILITY AND STABILITY OF MATHEMATICAL MODELS

Abderrahman Iggidr, INRIA (French National Institute for Research in Computer Science and Control), Ur-Lorraine and University of Metz, France, Abderrahman.Iggidr@inria.fr

Keywords: accessibility, asymptotic stability, attractivity, chemostat, closed-loop, controllability, differential equation, finite dimensional systems, input, Lie algebra, Lotka-Volterra systems, Lyapunov functions, nonlinear systems, observability, observer, output, stabilization, state.

Contents

Glossary

Bounded set: A subset B of $\mathbb{R}^n$ is bounded if there exists a positive real number M such that $\| x \| \leq M$ for all $x \in B$.

Open set: A subset E of $\mathbb{R}^n$ is open in $\mathbb{R}^n$ if for every point $p \in E$ there is a real number $\epsilon > 0$ such that the open ball $B(p, \epsilon) = \{x \in \mathbb{R}^n \ : \| x - p \| < \epsilon\}$ is contained in E.

Closed set: A subset F of $\mathbb{R}^n$ is closed in $\mathbb{R}^n$ if its complement $E = \mathbb{R}^n \setminus F$ is open in $\mathbb{R}^n$.

Closure of a set: The closure $\bar{A}$ of a set A is the smallest closed set containing A.

Compact set: A subset K of $\mathbb{R}^n$ is compact in $\mathbb{R}^n$ if it is closed and bounded.

Connected set: A subset Ω of $\mathbb{R}^n$ is said to be connected if it can not be represented as the union of two disjoint non-empty open sets of $\mathbb{R}^n$.

State vector: Vector whose elements are the state variables of a dynamical system.

State space: The space to which the state vector belongs.

Continuous time systems: Systems whose evolution is described by differential equations.

Discrete time systems: Systems whose evolution is described by difference equations.

Time invariant systems: Dynamical systems for which the time derivative of the state vector at a time t depends only on the state vector at the current time t and does not depend explicitly on time. For instance, the parameters of the model of a time-invariant system are constants.

Vector field: A vector field X on $\mathbb{R}^n$ is a function that associates a vector $X(x) \in \mathbb{R}^n$ to every point $x \in \mathbb{R}^n$. A vector field X is always associated to a differential equation $\dot{x} = X(x)$.

Characteristic polynomial: The characteristic polynomial of a square $n \times n$ matrix A is defined by $P_A(X) = det(A - XI_n)$ where I_n is the $n \times n$ identity matrix and $det(M)$ is the determinant of the square matrix M.

Eigenvalue: A scalar λ is an eigenvalue of a square matrix A if there exists a vector $v \neq 0$ such that $A.v = \lambda v$. The eigenvalues of a square matrix A are the roots of its characteristic polynomial.

Summary
This article presents an overview of three fundamental concepts in Mathematical System Theory: controllability, stability and observability. These properties play a prominent role in the study of mathematical models and in the understanding of their behavior. They constitute the main research subject in Control Theory. Historically the tools and techniques of Automatic Control have been developed for artificial engineering systems but nowadays they are more and more applied to "natural systems". The main objective of this article is to show how these tools can be helpful to model and to control a wide variety of natural systems.

1 Introduction

The main goal of this article is to develop in some details some notions of Control Theory introduced in the article **(Basic principles of mathematical modeling)**. It concerns more specifically three structural properties of control systems: controllability, stability and observability. Based on the references listed in the end of this article, we give a survey of these three properties with applications to various LSS examples.

By a control system we mean a dynamical system evolving in some state space and that can be controlled by the user. More precisely we are interested in the study of systems that can be modeled by differential (respectively difference) equations of the form

$$\begin{cases} \dot{x}(t) = \dfrac{dx}{dt} = X(x(t), u(t)), \\ y(t) = h(x(t)), \end{cases} \tag{1}$$

where the variable t represents the time, the vector $x(t)$ is the state of the system at time t, the vector $u(t)$ is the input or the control, i.e., the action of the user or of the environment and the vector $y(t)$ is the output of the system that is, the available information that can be measured or observed by the user. The dynamics function X indicates how the system changes over time.

We shall address the following problems:

- When the whole state $x(t)$ is not available for measurement, how is it possible to use the information provided by $y(t)$ together with the dynamics (1) in order to get a "good" estimation of the real state of the system? This turns out to be an observability problem.
- How to use the control $u(t)$ in order to meet some specified needs? This is the problem of controllability and stabilizability.

To illustrate these various concepts, we explain them through a simple LSS system: an epidemic model for the transmission of an infectious disease. An homogeneous population is divided into four classes S, E, I and T according to the heath of its individuals. Let $S(t)$ denote the number of individuals who are susceptible to the disease, i.e., who are not yet infected at time t. $E(t)$ denotes the number of members at time t who are exposed but not yet infected. $I(t)$ denotes the number of infected individuals, that is, who are infectious and able to spread the disease by contact with individuals who are susceptible. $T(t)$ is the number at time t of treated individuals. The total population size is denoted $N = S + E + I + T$. The dynamic of the disease can be described by the following differential system:

$$\begin{cases} \dfrac{dS}{dt} = bN - \mu S - \beta \dfrac{SI}{N}, \\ \dfrac{dE}{dt} = \beta \dfrac{SI}{N} - (\mu + \epsilon)E, \\ \dfrac{dI}{dt} = \epsilon E - (r + d + \mu)I, \\ \dfrac{dT}{dt} = rI - \mu T, \\ \dfrac{dN}{dt} = (b - \mu)N - dI, \end{cases} \tag{2}$$

where the parameter b is the rate for natural birth and μ that of natural death. The parameter

β is the transmission rate, d is the rate for disease-related death, ϵ is the rate at which the exposed individuals become infective and r is the per-capita treatment rate. The parameters b, μ, β, d and ϵ are assumed to be constant. The per-capita treatment rate may vary with time.

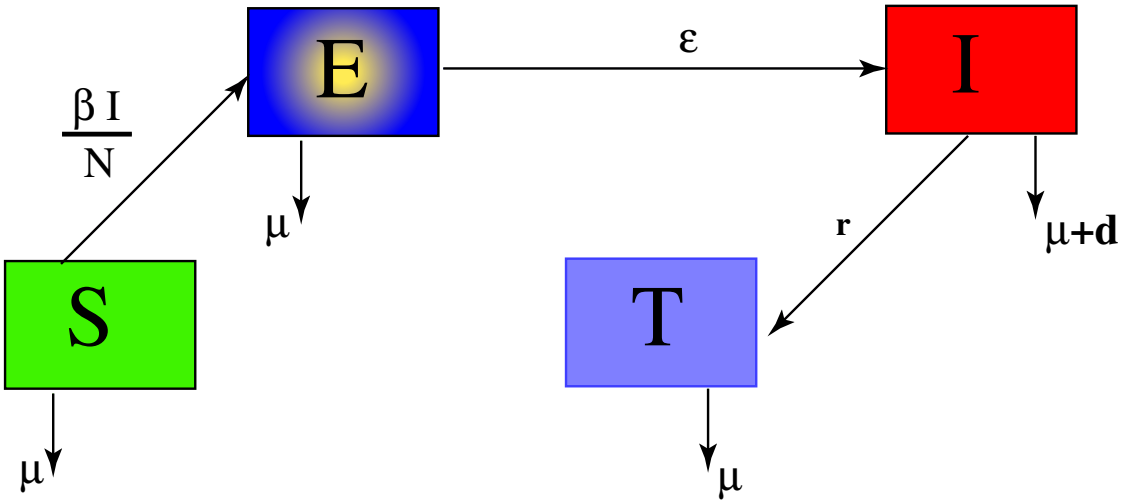

Figure 1: This diagram illustrates the dynamical transfer of the population.

The system (2) can be written by using the fractions $s = S/N$, $e = E/N$, $i = I/n$ and $\tau = T/N$ of the classes S, E, I and T in the population. These fractions satisfy the system of differential equations:

$$\begin{cases} \dfrac{ds}{dt} = b - bs - (\beta - d)si, \\ \dfrac{de}{dt} = \beta si - (b + \epsilon)e + dei, \\ \dfrac{di}{dt} = \epsilon e - (r + d + b)i + di^2, \\ \dfrac{d\tau}{dt} = ri - b\tau + di\tau. \end{cases} \tag{3}$$

Since $s + e + i + \tau = 1$, it is sufficient to consider the following system

$$\begin{cases} \dfrac{ds}{dt} = b - bs - (\beta - d)si, \\ \dfrac{de}{dt} = \beta si - (b + \epsilon)e + dei, \\ \dfrac{di}{dt} = \epsilon e - (r + d + b)i + di^2. \end{cases} \tag{4}$$

If we suppose that the proportion of infected individuals can be measured and that we can control the population by choosing the treatment rate then the above system can be seen as a control system of the form (1) with: the state of the system is $x(t) = (s(t), e(t), i(t))$, the control is $u(t) = r(t)$ and the measurable output is $y(t) = i(t)$. The state space is

$$\Omega = \{x \in \mathbb{R}^3 \ : \ 0 \leq s \leq 1, \ 0 \leq e \leq 1, \ 0 \leq i \leq 1, \ s + e + i \leq 1\}.$$

The aim of Control Theory is to give answers to the following questions:

1. Given two configurations x_1 and x_2 of the system (3), is it possible to steer the system from x_1 to x_2 by choosing an appropriate treatment strategy $r(t)$? For a given initial state $x_0 = (s_0, e_0, i_0)$ and a given time T, what are all the possible states that can be reached in time T by using all the possible control functions $r(t)$?
2. Does the system (3) possess equilibrium points? Are they stable or unstable? How can the control be chosen in order to make a given equilibrium (for instance, the disease-free equilibrium) globally asymptotically stable?
3. Since only some variables can be measured (for instance, the proportion $i(t)$ of infected individuals in the above epidemic example), is it possible to use them together with the dynamics (3) of the system in order to estimate the non measurable variables $s(t)$ and $e(t)$? and how this can be done?

The questions 1 involve the controllability of control systems. We shall give in Section 2 a general definition of this notion as well as some simple useful criteria that allow to study the controllability of some nonlinear systems. The questions 2 invoke the stability and the stabilization of nonlinear system, this will be the subject of Section 3 where some classical stability results are exposed. The questions 3 are connected to the observability and the construction of observers for dynamical systems. These problems will be addressed in Section 4 and Section 5.

In each section we shall apply the different tools to various models such as predator-prey systems, fisheries and bioreactors. Concerning the epidemic model (2-4), the divers problems mentioned above are still under investigation. We can give here just some partial answers. It can be shown that if the per-capita treatment rate satisfies

$$r \geq r_0 = \frac{\beta\,\epsilon}{\epsilon + b} - d - b \tag{5}$$

then the system has a unique equilibrium point in the domain Ω. This equilibrium is the disease free equilibrium $s = 1$, $e = i = \tau = 0$ and it is globally asymptotically stable, that is, if $(s(0), e(0), i(0)) \in \Omega$ is any initial condition then the solution of the system (4) starting from this state will converge to the disease free equilibrium, i.e., $s(t) \to 1$, $e(t) \to 0$ and $i(t) \to 0$ as $t \to \infty$. Therefore, if the treatment satisfy the condition (5) then the disease will die out. If $r < r_0$ then the disease free equilibrium becomes unstable and in this case, there is another equilibrium which belongs to the interior of the domain Ω. This equilibrium is the endemic equilibrium and it is globally asymptotically stable within the interior of Ω provided that $r < r_0$. The disease will be endemic in this case.

2 Controllability

In this section, we present an elementary overview of an important property of a system, namely that of controllability. This concept has been briefly introduced and explained in **(Basic principles of mathematical modeling)** . Here, we deal with the problem of testing controllability of systems of the type

$$\dot{x}(t) = X(x(t), u(t)). \tag{6}$$

The state vector $x(t)$ belongs to the state space M which will always be here $\mathbb{R}^n$ or an open connected subset of $\mathbb{R}^n$. The control functions $t \mapsto u(t)$ are defined on $[0,\ \infty)$ and take values

in a connected subset U of $\mathbb{R}^m$. These control functions are assumed to belong to an *admissible* control set $\mathcal{U}_{ad}$. This admissible control set is generally specified by the problem considered. Here it is assumed that $\mathcal{U}_{ad}$ contains the set of piecewise constant functions as a dense subset, i.e., any admissible control function can be "approximated" by piecewise constant functions. The dynamics function $X : M \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is assumed to be analytic on (x, u). For each control value $u \in U$, we denote by X^u the vector field ("vector function") defined by $X^u(x) = X(x, u)$ for all $x \in M$.

A real-valued function f defined on some open set $\mathcal{O} \subset \mathbb{R}$ is said to be analytic if for each $x_0 \in \mathcal{O}$, there exists a positive real number $r > 0$ such that $f(x)$ is the sum of a power series for all x with $|x - x_0| < r$:

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n \quad \text{for all } x \text{ satisfying } |x - x_0| < r.$$

For an analytic function, the coefficients a_n can be computed as $a_n = \frac{f^{(n)}(a)}{n!}$
In a similar way a function $f : \mathcal{V} \subset \mathbb{R}^k \to \mathbb{R}$ of several variables is said to be analytic if it is locally given by power series. A vector function

$$\begin{array}{lll} X: & \mathcal{V} \subset \mathbb{R}^k & \to \mathbb{R}^n \\ & x = (x_1, \ldots, x_k) & \mapsto X(x) = \begin{pmatrix} X_1(x_1, \ldots, x_k) \\ \vdots \\ X_n(x_1, \ldots, x_k) \end{pmatrix} \end{array}$$

is analytic if all its components X_i are analytic.

For each $x \in M$ and each control function $u(.) \in \mathcal{U}_{ad}$, we denote by $X_t^{u(.)}(x)$ the solution of (6) satisfying $X_0^{u(.)}(x) = x$. We assume moreover that $X_t^{u(.)}(x)$ is defined for all $t \in [0, \infty)$. For instance, if X is a linear vector function, that is, $X(x, u) = Ax + Bu$, with A and B being matrices with appropriate dimensions, then $X_t^{u(.)}(x) = e^{tA}x + \int_0^t e^{(t-s)A}Bu(s)ds$.

2.1 What is controllability

A state $x_1 \in M$ is *reachable* from $x_0 \in M$ at time $T \geq 0$ if there exists a control $u(.) \in \mathcal{U}_{ad}$ such that $x_1 = X_T^{u(.)}(x_0)$. The set of reachable states from x_0 at time T will be denoted $\mathcal{A}^T(x_0)$ and the reachable set from x_0 will be denoted $\mathcal{A}(x_0) = \bigcup_{T \geq 0} \mathcal{A}^T(x_0)$.
For the class of piecewise constant controls, the reachable set at time T from x_0 is the set of points of M of the form:

$$x = X_{t_p}^{u_p} o \ldots o X_{t_2}^{u_2} o X_{t_1}^{u_1}(x_0), \ u_1, \ldots, u_p \in U, \ t_1 + \ldots + t_p = T.$$

This formula means that the constant control u_i is applied during time t_i, i.e., $u(t) = u_1$ for $t \in [0, \ t_1)$, $u(t) = u_2$ for $t \in [t_1, \ t_1 + t_2)$, $\ldots$, $u(t) = u_i$ for $t \in [t_1 + \ldots t_{i-1}, \ t_1 + \ldots t_{i-1} + t_i)$.

The system (6) will be said to be *accessible from* x_0 if the reachable set $\mathcal{A}(x_0)$ has a nonempty interior in the state space M. When the set $\mathcal{A}(x_0)$ is equal to the whole state space M then the system is *controllable from* x_0 and it is *controllable* (or completely controllable) if this property holds for any $x_0 \in M$. More precisely, the system (6) is controllable if any point of M is reachable from any other point of M, i.e., for any two states $x_0, \ x_1 \in M$ there exist a finite time T (that may depend on $(x_0, \ x_1)$) and an admissible control function $u(.) : [0, T] \to U$

such that $x_1 = X_T^{u(.)}(x_0)$.
The system (6) is *strongly controllable* if for any given time $T > 0$ any point of the state space M is reachable from any other point of M in T or fewer units of time.

Before giving some criteria for accessibility and controllability of control systems, we give a simple LSS example to illustrate these notions.

Example: Let us consider a simple Lotka-Volterra system that models a population of a harvested prey in the presence of a predator

$$\begin{cases} \dot{N} = aN - bNP - q\ uN \\ \dot{P} = -cP + ebNP \end{cases} \tag{7}$$

In these equations N and P represent, respectively, the prey and the predator populations, a is the prey growth rate, b is the predator attack rate, c is the predator mortality rate, and e is the conversion efficiency of predators. The parameter q represents the catchability coefficient of the prey and u is the harvesting effort. The term quN is the rate of harvest when the effort is u and the prey stock is N. All the variables and the constants are positive. Here the state is $x = (N, P)$, the state space is the positive quadrant $M = \mathbb{R}^2_+ = \{(N, P) \in \mathbb{R}^2 \ : \ N > 0,\ P > 0\}$, the effort u represents the harvesting policy and can be seen as an input or a control term. We suppose that u can take two values 0 or 1. Therefore our admissible control set $\mathcal{U}_{ad}$ is simply the set of piecewise constant functions defined on $[0,\ \infty)$ with values in $U = \{0, 1\}$. The dynamics vector function is $X(x, u) = \begin{pmatrix} aN - bNP - q\ uN \\ -cP + ebNP \end{pmatrix}$. The values of the control u generate two vector fields $X^0(x) = \begin{pmatrix} aN - bNP \\ -cP + ebNP \end{pmatrix}$ and $X^1(x) = \begin{pmatrix} aN - bNP - qN \\ -cP + ebNP \end{pmatrix}$. The trajectory of a vector field X through $x \in M$ is the curve $\{X_t(x),\ t \in [0,\ \infty)\}$ where $X_t(x)$ is the solution of $\dot{x} = X(x)$ with initial condition $X_0(x) = x$ (we suppose that the solution is defined for all $t \geq 0$). Here, the trajectories of the vector fields X^0 and X^1 in the interior of the positive quadrant are closed curves (see Figure 2). More precisely the trajectories are the level curves of, respectively, the real-valued functions

$$V_0(N, P) = e\,b\,N - c\ \log(\tfrac{e\,b}{c}N) + b\,P - a\ \log(\tfrac{b}{a}P) - (a + c)$$
$$V_1(N, P) = e\,b\,N - c\ \log(\tfrac{e\,b}{c}N) + b\,P - (a - q)\ \log(\tfrac{b}{a-q}P) - (a - q + c).$$

The horizontal and the vertical axes are particular trajectories for both X^0 and X^1.

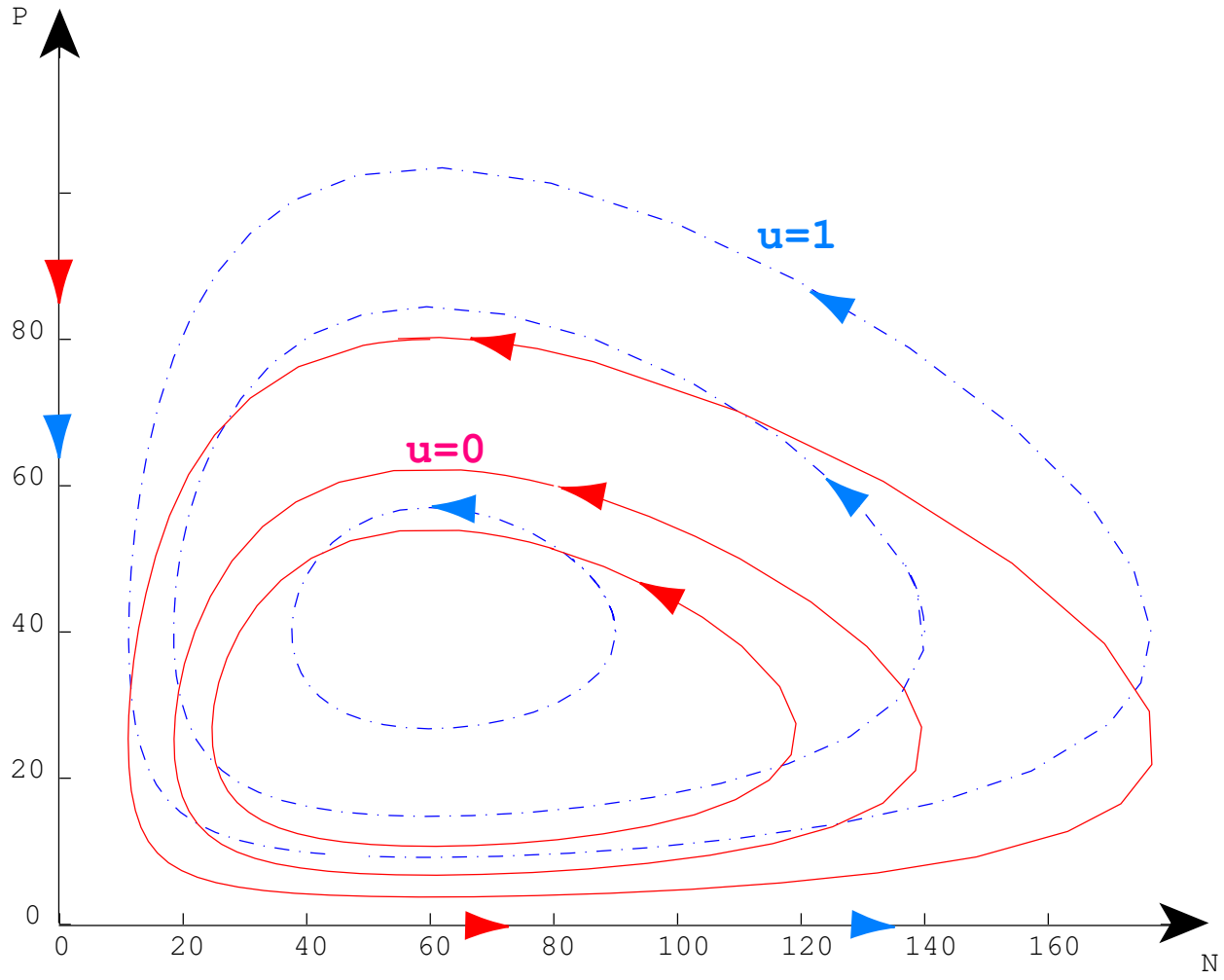

Figure 2: Trajectories of X^0 and X^1 for $a = 0.4$, $b = 0.01$, $c = 0.3$, $e = 0.5$, $q = 0.15$.

Now, given an initial state x_0, what is the reachable set from x_0? if x_0 is of the form $x_0 = (N_0, 0)$ then the reachable set is $\mathcal{A}(x_0) = \{(N, 0) \ : \ N \geq N_0\}$ and if $x_0 = (0, P_0)$ then $\mathcal{A}(x_0) = \{(0, P) \ : \ P \leq P_0\}$ since the vertical and the horizontal axes are positively invariant sets for the system (7). The interior M of the positive quadrant is positively invariant. Therefore if x_0 belongs to M then the whole trajectory starting from x_0 is contained in M and so $\mathcal{A}(x_0) \subset M$. The reachable set of such state x_0 is actually the whole set M. For, let x_1 be any point of M, we shall show that $x_1 \in \mathcal{A}(x_0)$. There exist two positive numbers r_0 and r_1 such that the point x_0 belongs to the level curve $\mathcal{C}_{r_0}$ of V_0 defined by $V_0(x) = r_0$ and the point x_1 belongs to the set $\mathcal{C}_{r_1}$: $V_0(x) = r_1$. It can be proved that along each level curve $\mathcal{C}_r$ of X^0, there is a point where the vector field X^1 points inward $\mathcal{C}_r$ and a point where the vector field X^1 points outward $\mathcal{C}_r$. This implies that the control $u = 1$ allows to pass from a given level set $\mathcal{C}_r$ to a "smaller" one as well as to a "larger" one. Therefore one can go from the level curve $\mathcal{C}_{r_0}$ to the level curve $\mathcal{C}_{r_1}$ by using a piecewise constant control $u = 0$ or $u = 1$. Hence the system can be driven from the state x_0 to the state x_1. The system is then completely controllable on the state space M. It must be noticed that there are many other possible strategies to achieve the same goal, for instance, the same reasoning can be done by considering the level curves of X^1 instead of those

of X^0. The figure 3 illustrates possible control schemes to go from x_0 to x_1.

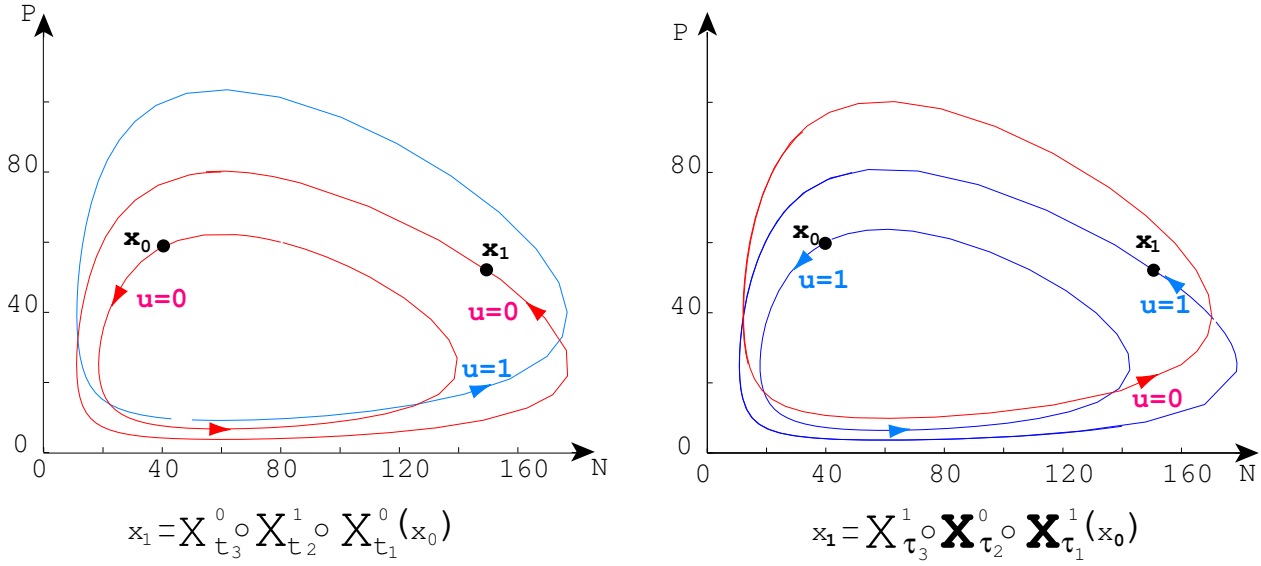

Figure 3: Two control strategies to drive the system from the state x_0 to the state x_1.

2.2 Controllability of linear systems

We have seen that the above system is controllable on the positive quadrant. To show this we have used the geometrical properties of the trajectories of the vector fields generated by the different values of the control. Unfortunately, it is not often possible to know explicitly the phase portrait of a vector field. So there is a need to have calculable criteria for testing controllability or at least accessibility of a given system. For linear systems, there is a simple controllability criterion known as *Kalman's controllability rank condition*. A linear system is a system governed by

$$\begin{cases} \dot{x} = X(x,u) = Ax + Bu, \\ x \in \mathbb{R}^n,\ u \in U \subset \mathbb{R}^m, \end{cases} \tag{8}$$

where A and B are respectively $n \times n$ and $n \times m$ matrices. We start by exploring the controllability properties when the control set is $U = \mathbb{R}^m$, i.e., there are no restrictions on the size of controls. It can be proved that the following are equivalent: the linear system (8) is **(i)** controllable, **(ii)** accessible from any point $x_0 \in \mathbb{R}^n$, **(iii)** accessible from the origin $x_0 = 0$. The reachable set from the origin $\mathcal{A}(0)$ is a linear subspace of $\mathbb{R}^n$. More precisely, it is the image of the linear map:

$$\begin{array}{rcl} \mathbb{R}^{m+n} & \longrightarrow & \mathbb{R}^n \\ (u_1,\ldots,u_n) & \longmapsto & (B, AB, A^2B, \ldots, A^{n-1}B) \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^{n} A^{i-1}Bu_i\ . \end{array}$$

The $n \times nm$ block matrix $R(A,B) = (B, AB, A^2B, \ldots, A^{n-1}B)$ whose columns are the columns of B, AB, $\ldots$, $A^{n-1}B$ is called the *Kalman controllability matrix*. The controllability of the linear system (8) is related to the rank of this matrix as follows

Theorem 2.1 *The system (8) is controllable if and only if the $n \times nm$ Kalman controllability matrix*

$$R(A,B) = \left(B, AB, A^2B, \ldots, A^{n-1}B\right)$$

is of rank n (the dimension of the state space).

In this case we also say that the pair (A, B) is controllable. It must be noticed that for a controllable linear system, every state x_1 can be reached from any state x_0 in a time interval of arbitrary length (i.e., as small as one wants) provided that the control values set U is unbounded. For linear systems with unbounded control values set, there is an equivalence between controllability and strong controllability.

When the linear system (8) is not controllable, we have $rank(R(A,B)) = r < n$. Let $c_1, c_2, \ldots, c_r$ be r linearly independent columns of the matrix $R(A,B)$ and let $e_1, \ldots, e_{n-r}$ be any vectors of $\mathbb{R}^n$ such that $\{c_1, c_2, \ldots, c_r, e_1, \ldots, e_{n-r}\}$ is a basis of $\mathbb{R}^n$. Then the matrix $T = (c_1, c_2, \ldots, c_r, e_1, \ldots, e_{n-r})$ is an $n \times n$ invertible matrix and with the linear change of coordinates $x = Tz$, the system (8) is transformed into

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + A_{12}z_2 + B_1u, \\ \dot{z}_2 = A_{22}z_2, \\ z_1 \in \mathbb{R}^r, \ z_2 \in \mathbb{R}^{n-r} \end{cases} \tag{9}$$

where the pair (A_{11}, B_1) is a controllable pair of matrices. We can remark that the control cannot act on the z_2 component of the state. The eigenvalues (or the eigenvectors) of the matrix A_{22} are called the "uncontrollable modes" of the system.

Controllability with bounded controls: The Kalman's rank condition is a necessary condition for controllability of linear systems. It is a sufficient condition when the control values set U is unbounded. When there is a bound on the magnitude of the control u, the rank condition is no more sufficient to get controllability as it is illustrated by the following linear system

$$\dot{x} = x + u, \ x \in \mathbb{R}, \ u \in \mathbb{R}. \tag{10}$$

The rank controllability condition is obviously satisfied for this system. hence it is controllable and even strongly controllable if the control u can take any real value. Now, suppose that the admissible controls are bounded by a positive constant K, that is, for all $u(.) \in \mathcal{U}_{ad}$, one has $|u(t)| \leq K$ for all positive time t. The solution of (10) is

$$x(t) = e^t \left(x(0) + \int_0^t e^{-s}u(s)ds \right).$$

Since $|u(s)| \leq K$, we have $\left| \int_0^t e^{-s}u(s)ds \right| \leq K$. It follows that $x(t) > K$ for all positive time t if $x(0) > 2K$, and therefore the system is not controllable on $\mathbb{R}$.

The above example shows that controllability of a linear system cannot in general be expected if there are restrictions on the size of controls. There is actually an additional necessary condition for controllability when the set U is bounded:

Theorem 2.2 *a. Suppose that the control set U is bounded. Then a necessary condition for controllability of the linear system (8) is that every eigenvalue of A have its real part equal to zero.*
b. If this necessary condition is satisfied and the set U is a neighborhood of the origin in $\mathbb{R}^m$, then the linear control system with controls in U is controllable whenever the Kalman's rank condition holds, i.e., rank $(B, AB, A^2B, \ldots, A^{n-1}B) = n$.

It is interesting to remark that

- the controllability of linear systems with unbounded controls is a "generic" property of linear systems. This means, roughly speaking, that almost all the pairs (A, B) are controllable when $U = \mathbb{R}^m$. More precisely, if a given pair (A, B) does not satisfy the rank condition then by a "small perturbation" of the entries of the matrices, the pair becomes controllable;
- linear systems are generically not controllable with bounded control because the necessary condition is easily violated when the entries of the matrix A are perturbed.

2.3 Accessibility criteria for nonlinear systems

For each control value $u \in U$, let $X^u = X(., u)$ be the vector function (vector field) corresponding to the constant control u. We then have a family $\mathcal{F} = \{X^u, u \in U\}$ of vector fields parametrized by the values of the control. To get an accessibility criterion for the system (6) analogous to the Kalman's rank condition for linear systems we need to introduce some new vector fields generated by the elements of $\mathcal{F}$. For any vector functions X and Y, it is possible to define a new vector function $[X, Y]$, called the *Lie bracket* of X and Y, and defined by

$$\forall x \in \mathbb{R}^n,\ [X, Y](x) = \frac{\partial Y}{\partial x}(x)X(x) - \frac{\partial X}{\partial x}(x)Y(x)$$

where $\frac{\partial Y}{\partial x}(x)$ is the Jacobian of the vector function evaluated at the point x. In a coordinates system $(x_1, x_2, \ldots, x_n)$, if

$$X(x) = \begin{pmatrix} X_1(x_1, x_2, \ldots, x_n) \\ X_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ X_n(x_1, x_2, \ldots, x_n) \end{pmatrix} \text{ and } Y(x) = \begin{pmatrix} Y_1(x_1, x_2, \ldots, x_n) \\ Y_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ Y_n(x_1, x_2, \ldots, x_n) \end{pmatrix}$$

then,

$$[X, Y](x) = \begin{pmatrix} Z_1(x_1, x_2, \ldots, x_n) \\ Z_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ Z_n(x_1, x_2, \ldots, x_n) \end{pmatrix} \text{ with } Z_i(x) = \sum_{j=1}^{j=n} \frac{\partial Y_i}{\partial x_j}(x)X_j(x) - \frac{\partial X_i}{\partial x_j}(x)Y_j(x).$$

For example, let X be a linear vector function, say, $X(x) = Ax$ with A an $n \times n$ matrix and let Y be a constant vector function, that is, $Y(x) = b \in \mathbb{R}^n$. Then $[X, Y]$ is a constant vector function and $[X, Y](x) = -A\, b$, for all $x \in \mathbb{R}^n$.

A geometric interpretation of the Lie bracket: Let us consider a control system

$$\begin{cases} \dot{x} = u_1 X_1(x) + u_2 X_2(x) \\ x \in \mathbb{R}^n, \ (u_1, u_2) \in \mathbb{R}^2. \end{cases} \tag{11}$$

Let $x_0 \in \mathbb{R}^n$ be a given state such that the vectors $X_1(x_0)$ and $X_2(x_0)$ are linearly independent, and let $E(x_0) = span\{X_1(x_0), X_2(x_0)\}$ be the vector subspace of $\mathbb{R}^n$ spanned by the vectors $X_1(x_0)$ and $X_2(x_0)$. From the point x_0, we can steer in all directions contained in $E(x_0)$ by using appropriate constant controls. Is it possible to steer in another direction that does not belong to $E(x_0)$? If the answer is no then the system can never be accessible from x_0. It turns out that one also can steer in the direction of $[X_1, X_2](x_0)$ by choosing appropriate piecewise constant control functions. So, if $[X_1, X_2](x_0)$ does not belong to $E(x_0)$, then one has a third steering direction. This shows that the Lie bracket plays a prominent role in determining the reachable set from a given state x_0.

To the control system (6) and the generated family of vector functions $\mathcal{F}$ we associate a set of vector functions called the *accessibility Lie algebra* of the system (6) and denoted $\mathcal{L}ie(\mathcal{F})$. The set $\mathcal{L}ie(\mathcal{F})$ is the Lie algebra of vector functions generated by the family $\mathcal{F}$: every element of $\mathcal{L}ie(\mathcal{F})$ is a combination of repeated Lie brackets of the form

$$[X^{u^k}, [X^{u^{k-1}}, [\ldots, [X^{u^2}, X^{u^1}] \ldots]]], \text{ with } u^i \in U, \ X^{u^i} \in \mathcal{F}.$$

For instance, if u^1, u^2 are any two control values then the vector functions

$$[X^{u^1}, X^{u^2}], \ [X^{u^1}, [X^{u^1}, X^{u^2}]], \ [X^{u^2}, [X^{u^1}, [X^{u^1}, X^{u^2}]]],$$

are in $\mathcal{L}ie(\mathcal{F})$.

The control system (6) satisfies the *accessibility rank condition* at the point x_0 if the set

$$\mathcal{L}ie(\mathcal{F})(x_0) = \{X(x_0), \ X \in \mathcal{L}ie(\mathcal{F})\}$$

is a vector space of dimension n (the dimension of the state space). This rank condition is analogous to the Kalman's rank condition. It allows to get the following accessibility criterion for analytic nonlinear systems:

Theorem 2.3 *The analytic control system (6) is accessible from the point x_0 if and only if the accessibility rank condition holds at x_0.*

Remarks: 1. If the control system is not analytic but only smooth (C^∞), then the rank condition is still sufficient but not necessary.
2. For the linear system (8) with $U = \mathbb{R}^m$, it is easy to see that the elements of $\mathcal{L}ie(\mathcal{F})(x_0)$ are generated by elements of the form $A^k Bu$, $k = 0, \ldots, n-1$, or $Ax_0 + Bu$, where $u \in \mathbb{R}^m$. Therefore, for the linear system (8), the accessibility rank condition at the origin reduces to the Kalman controllability rank condition.

Example: Consider the following system defined on $\mathbb{R}^3$:

$$\dot{x} = u_1 \underbrace{\begin{pmatrix} -2x_3e^{x_1} \\ e^{x_1} \\ 0 \end{pmatrix}}_{X_1(x)} + u_2 \underbrace{\begin{pmatrix} -2(x_2+1) \\ 0 \\ 1 \end{pmatrix}}_{X_2(x)}, \quad x = (x_1, x_2, x_3), \quad (u_1, u_2) \in \mathbb{R}^2. \tag{12}$$

For any state $x \in \mathbb{R}^3$, the vectors $X_1(x)$ and $X_2(x)$ are linearly independent but $[X_1, X_2](x) = 2(x_2+1)X_1(x)$. Since $[X_1, X_2]$ is collinear with X_1, $\mathcal{L}ie(X_1, X_2)$ has a constant dimension equal to 2. Therefore, since the system (12) is analytic and the rank condition fails, it is is not accessible from any point $x \in \mathbb{R}^3$, that is, the reachable set $\mathcal{A}(x)$ has an empty interior whatever what the control is.

This fact can be seen directly by performing a change of coordinates that allows to rewrite the system (12) in a simpler form. Let $\Phi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ the map defined by $x = (x_1, x_2, x_3) \mapsto \Phi(x) = y = (y_1, y_2, y_3)$ with $y_1 = x_1 + 2x_2x_3 + 2x_3$, $y_2 = x_2$ and $y_3 = x_3$. The map Φ is bijective and its inverse is given by: $\Phi^{-1}(y_1, y_2, y_3) = (y_1 - 2y_2y_3 - 2y_3)$. It is clear that the maps Φ and Φ^{-1} are continuously differentiable. Therefore Φ is a global diffeomorphism and allows to define a new coordinate system $y = \Phi(x)$. We also make the inputs transformation $v_1 = \exp(y_1 - 2y_2y_3 - 2y_3)u_1$, $v_2 = u_2$. Hence in the new coordinates, the system (12) is expressed by:

$$\begin{cases} \dot{y}_1 &= 0 \\ \dot{y}_2 &= v_1 \\ \dot{y}_3 &= v_2 \end{cases}$$

It is clear that the reachable set from any state $x \in \mathbb{R}^3$ is contained in a vector space of dimension 2 and, hence, it is of empty interior. ■

Control affine systems: A control affine system is a control system of the form

$$\begin{cases} \dot{x} = X_0(x) + u_1X_1(x) + \ldots + u_mX_m(x), \\ x \in M, \ u = (u_1, \ldots, u_m) \in U \subset \mathbb{R}^m, \end{cases} \tag{13}$$

where $X_0, X_1, \ldots, X_m$ are analytic vector fields on the state space M, and $u_1, \ldots, u_m$ are the control functions. The vector field X_0 is called the drift and the vector fields $X_1, \ldots, X_m$ are called controlled vector fields. We assume that the control values set U contains m linearly independent points of $\mathbb{R}^m$. For example if $m = 2$ then this assumption is fulfilled if U contains the two points $(1, 0)$ and $(0, 1)$. The interest of this assumption is that the Lie algebra generated by the family $\mathcal{F}$ corresponding to the control system (13) is independent of U and it is equal to the Lie algebra generated by the vector fields $X_0, X_1, \ldots, X_m$, that is,

$$\mathcal{L}ie(\mathcal{F}) = \mathcal{L}ie\{X_0, X_1, \ldots, X_m\}.$$

For control affine systems (13), with U satisfying the above assumption, a controllability criterion is available:

Theorem 2.4 *Assume that* $dim(\mathcal{L}ie\{X_1, \ldots, X_m\}(x)) = n = dimM$ *for all* $x \in M$. *Then*

(i) *the control affine system (13) is controllable whenever there are no restrictions on the size of controls,*

(ii) *the driftless ($X_0 = 0$) system (13) is controllable even if there are restrictions on the size of controls, provided that the convex hull of the constraint set U is a neighborhood of the origin in $\mathbb{R}^m$.*

Example: Consider the following simplified model of maneuvering a car:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = u_1 \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{X_1(x)} + u_2 \underbrace{\begin{pmatrix} \sin x_3 \\ \cos x_3 \\ 0 \end{pmatrix}}_{X_2(x)}, \quad x = (x_1, x_2, x_3) \in \mathbb{R}^3, \quad (u_1, u_2) \in U \subset \mathbb{R}^2. \tag{14}$$

The center of the front axle has coordinates $(x_1, x_2) \in \mathbb{R}^2$, while the rotation of this axle is given by the angle x_3. The controls are the steering wheel moves u_1 and the engine speed u_2. This system is a driftless control affine system. It is reasonable to suppose that the controls take values in a bounded neighborhood of the origin in $\mathbb{R}^2$. We have $dim(\mathcal{L}ie\{X_1, X_2\}(x)) = 3$ for all $x \in \mathbb{R}^3$ since $[X_1, X_2](x) = \begin{pmatrix} -\cos x_3 \\ \sin x_3 \\ 0 \end{pmatrix}$ and for all $x \in \mathbb{R}^3$, $rank\{X_1(x), X_2(x), [X_1, X_2](x)\} = 3$. Therefore, Theorem 2.4 applies and hence the system is controllable. ■

Remark: It must be noticed that the sufficient controllability rank condition of the above theorem 2.4 invokes the Lie algebra generated only by the controlled vector fields $X_1, \ldots, X_m$. This is a stronger condition than the accessibility rank condition that invokes the Lie algebra $\mathcal{L}ie\{X_0, X_1, \ldots, X_m\}$. A control affine system may satisfy the accessibility rank condition without being controllable as illustrated by the following two-dimensional example:

$$\begin{cases} \dot{x}_1 = u, \\ \dot{x}_2 = x_1^2, \\ (x_1, x_2) \in M = \mathbb{R}^2, \ u \in U = \mathbb{R}. \end{cases} \tag{15}$$

Here, The drift vector field is $X_0 = \begin{pmatrix} 0 \\ x_1^2 \end{pmatrix}$, and the controlled vector field is $X_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. It is obvious that for all $x \in \mathbb{R}^2$, $dim(\mathcal{L}ie\{X_1\}(x)) = 1$. We have $[X_0, X_1] = \begin{pmatrix} 0 \\ 2x_1 \end{pmatrix}$, and $[X_1, [X_0, X_1]] = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$. Therefore, $dim(\mathcal{L}ie\{X_0, X_1\}(x)) = 2 = dim M$ everywhere, and hence, according to Theorem 2.3, the system is accessible. However, it is not controllable since the reachable set from a given point $x_0 = (x_{1_0}, x_{2_0})$ is contained in the half-space $x_2 \geq x_{2_0}$ (see Figure 4).

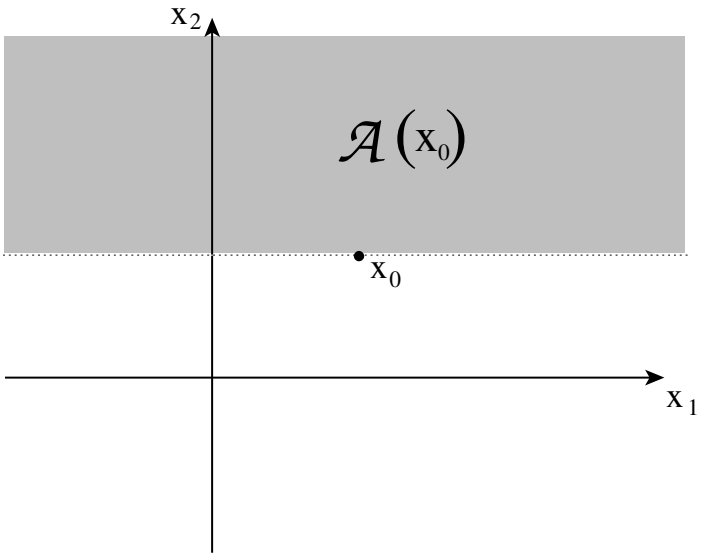

Figure 4: Reachable set from the state x_0 for the system (15).

Theorem 2.4 says that controllability with bounded controls is possible for driftless control affine systems, i.e., $X_0 = 0$. There also exists a controllability criterion with bounded controls for the system (13) in the presence of a drift vector field X_0 which is not equal to zero but which is *positively Poisson stable*. To simplify matters, we assume that the state space M is an open connected subset of $\mathbb{R}^n$. Let X be an analytic vector field on M and let for each point $x \in M$, $X_t(x)$ be the solution of the differential equation $\dot{x} = X(x)$ that satisfies $X_0(x) = x$. We assume that the vector field X is *complete*, which means that for any point $x \in M$, the solution $X_t(x)$ is defined for all $t \in [0, \infty)$. A point $x \in M$ is said to be *positively Poisson stable* for X if

$$\forall \epsilon > 0,\ \forall T > 0,\ \exists t \geq T :\ \|X_t(x) - x\| < \epsilon.$$

The vector field X is called positively Poisson stable if the set of positively Poisson stable points for X is dense in M. A set $N \subset M$ is dense in M if for any point $p \in M$ and any number $\epsilon > 0$, there exists $q \in N$ such that $\|p - q\| < \epsilon$. This means "near any point of M, there are points of N".

The linear vector field $X(x) = Ax$ with $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is a positively Poisson stable vector field on $\mathbb{R}^2$. Another example of positively Poisson stable vector field is given by the vector field X^0 corresponding to $u = 0$ in the predator-prey system (7).

Theorem 2.5 *Assume that the drift vector field X_0 is positively Poisson stable. Then the system (13) with controls constrained by $u_i \in \{-a_i, a_i\}$, $a_i > 0$, $i = 1 \dots m$ is controllable on M if for each point $x \in M$, one has dim $(\mathcal{L}ie\{X_0, X_1, \dots, X_m\}(x)) = n = dim M$.*

Remark: The control constraint U can actually be any subset of $\mathbb{R}^m$ whose convex hull contains the origin in its interior. For example, if $m = 2$ then the above theorem can be applied with $U = \{(-\frac{1}{2}, 0), (3, 0), (0, -1), (0, \frac{1}{4})\}$.

We go back to the harvested prey system (7) which can be written $\dot{x} = X_0(x) + uX_1(x)$, with $x = (N, P) \in M = \{(N, P) \in \mathbb{R}^2 : \ N > 0, P > 0\}$, $X_0 = \begin{pmatrix} aN - bNP \\ -cP + ebNP \end{pmatrix}$, and $X_1 = \begin{pmatrix} -qN \\ 0 \end{pmatrix}$. Let us evaluate the rank of the Lie algebra generated by the vector fields X_0 and X_1: The vectors $X_0(x)$ and $X_1(x)$ are linearly independent everywhere (in M) except at the points of the vertical line $N = \frac{c}{eb}$ so the vector space spanned by those two vectors is not of dimension 2 at every point $x \in M$. Hence, we have to compute their Lie bracket $[X_0, X_1](x) = \begin{pmatrix} 0 \\ ebNP \end{pmatrix}$. It is clear that the vector space spanned by the three vectors $X_0(x)$, $X_1(x)$ and $[X_0, X_1](x)$ is of dimension 2 at any point $x \in M$. Therefore, we have $dim\ (\mathcal{L}ie\{X_0, X_1\}(x)) = 2 = dimM$ and since the vector field X_0 is positively Poisson stable (its trajectories are closed curves in M), we can apply Theorem 2.5, that is, the system (7) is controllable on M with control constrained by, for instance, $u \in \{-1, 1\}$. However, we have seen that the system (7) can be controlled with a piecewise control that can take only the values 0 and 1. This fact has been established by using a qualitative geometrical reasoning. In this situation, the control values set is $U = \{0, 1\}$ and its convex hull is $co(U) = [0,\ 1]$ which does not contain zero in its interior so Theorem 2.5 can not apply. Below, we give another result that can be helpful to study controllability even if the system is not control affine nor the origin is an interior point of the convex hull of the control values set U.

Theorem 2.6 *Consider the analytic control system (6). Assume that*
(i) *the system is symmetric, i.e., $X(x, -u) = -X(x, u)$ for each $x \in M$ and each $u \in U$, or*
(ii) *each vector field of the family $\mathcal{F} = \{X^u, u \in U\}$ is positively Poisson stable.*
Then the system (6) is controllable if and only if the accessibility rank condition holds:

$rank(\mathcal{L}ie(\mathcal{F})(x)) = n = dimM,\ $ *for all* $x \in M$.

Let us consider again the predator-prey system (7):

$$\begin{cases} \dot{N} = aN - bNP - q\ uN, \\ \dot{P} = -cP + ebNP, \\ x = (N, P),\ N > 0,\ P > 0,\ u \in \{0, 1\}. \end{cases} \tag{16}$$

For this system we have $U = \{0, 1\}$ and $\mathcal{F} = \{X^u, u \in U\} = \{X^0, X^1\}$ with

$$X^0(x) = \begin{pmatrix} aN - bNP \\ -cP + ebNP \end{pmatrix},\ X^1(x) = \begin{pmatrix} aN - bNP - qN \\ -cP + ebNP \end{pmatrix}.$$

It is important to note that X^1 is not the controlled vector field X_1 (here, $X^0 = X_0$ and $X^1 = X_0 + X_1$). The two vector fields X^0 and X^1 are positively Poisson stable on M because, as we have seen before, their trajectories are the closed level curves of, respectively, the real-valued functions

$$V_0(N, P) = e\,b\,N - c\ \log(\tfrac{eb}{c}N) + b\,P - a\ \log(\tfrac{b}{a}P) - (a + c)$$
$$V_1(N, P) = e\,b\,N - c\ \log(\tfrac{eb}{c}N) + b\,P - (a - q)\ \log(\tfrac{b}{a-q}P) - (a - q + c).$$

Therefore, by Theorem 2.6, system (16) is controllable if and only if $rank(\mathcal{L}ie\{X^0, X^1\}(x)) = 2$ for all $x \in M$. This rank condition holds since

$$[X^0, X^1](x) = \begin{pmatrix} 0 \\ ebqNP \end{pmatrix}$$

and $rank\{X^0(x), X^1(x), [X^0, X^1](x)\} = 2$ for all $x \in M$.
It is worthwhile noticing that the system (16) remains controllable if $U = \{u_1, u_2\}$ for any non-negative real numbers $u_1 \neq u_2$. This results from the fact that the generated vector fields X^u are positively Poisson stable for any value of u and that the generated Lie algebra satisfies the rank condition provided $u_1 \neq u_2$.

3 Stability

3.1 General definitions

Let us consider a system of ordinary differential equations

$$\dot{x} = X(x) \ , \quad x \in \Omega \ , \tag{17}$$

where Ω is an open connected subset of $\mathbf{R}^n$ and X is a locally Lipschitz continuous map from Ω to $\mathbf{R}^n$. For each $x \in \Omega$, let us denote by $X_t(x)$ the solution of (17) satisfying $X_0(x) = x$.

An *equilibrium point* or a steady state is a state $x^e \in \Omega$ satisfying $X(x^e) = 0$. Corresponding to each equilibrium point x^e, we have a constant solution $X_t(x^e) \equiv x^e$ of (17).

Let $x^e \in \Omega$ be an equilibrium point. The system (17) is *stable* (we also say *Lyapunov stable*) at x^e or x^e is a *stable* equilibrium position for (17), if for each $\epsilon > 0$ there exists a positive real number δ such that for each x with $\|x - x^e\| < \delta$, the solution $X_t(x)$ is defined for all $t \geq 0$ and satisfies $\|X_t(x) - x^e\| < \epsilon$ for all $t > 0$. When (17) is not Lyapunov stable at x^e, we say that it is *unstable* at x^e, or that x^e is an unstable equilibrium for the system (17).

Lyapunov stability of an equilibrium x^e means that all solutions starting at nearby points stay nearby. The Lyapunov stability is an important property. For, let x^e be a desired steady-state of our system. Unpredictable perturbations may cause the system to deviate from x^e. Lyapunov stability guarantees that every state value taken by the system in its future evolution is not too far from the desired one if the perturbations are small.

The steady state x^e is said to be *attractive* (We also say that (17) is attractive at x^e) if there exists a neighborhood $\mathcal{U} \subset \Omega$ of x^e such that for any initial condition x belonging to $\mathcal{U}$, the corresponding solution $X_t(x)$ of (17) is defined for all $t \geq 0$ and tends to x^e as t tends to infinity, i.e., $\lim_{t \to +\infty} X_t(x) = x^e$.

Let $\mathcal{A}$ be the set of points $x \in \Omega$ such that

$$\lim_{t \to +\infty} X_t(x) = x^e$$

holds for all solutions $X_t(x)$ starting from x. Then $\mathcal{A}$ is connected and it is called the *region of attraction* of x^e.

The equilibrium x^e is globally attractive if $\mathcal{A} = \Omega$, i.e.,

$$\lim_{t \to +\infty} X_t(x) = x^e, \ \forall x \in \Omega.$$

Remark: An equilibrium can be attractive without being stable. This can be illustrated by the following classical two-dimensional system written in polar cordinates (r, θ):

$$\begin{cases} \dot{r} = r(1-r) \\ \dot{\theta} = \sin^2 \frac{\theta}{2}. \end{cases}$$

This system has two equilibria: the origin 0 of $\mathbb{R}^2$ and the point P defined by $r = 1$ and $\theta = 0$. It can be proved that P is attractive and that its attraction region is $\mathcal{A}(P) = \mathbb{R}^2 \setminus \{0\}$. However it is not stable. (see figure 5)

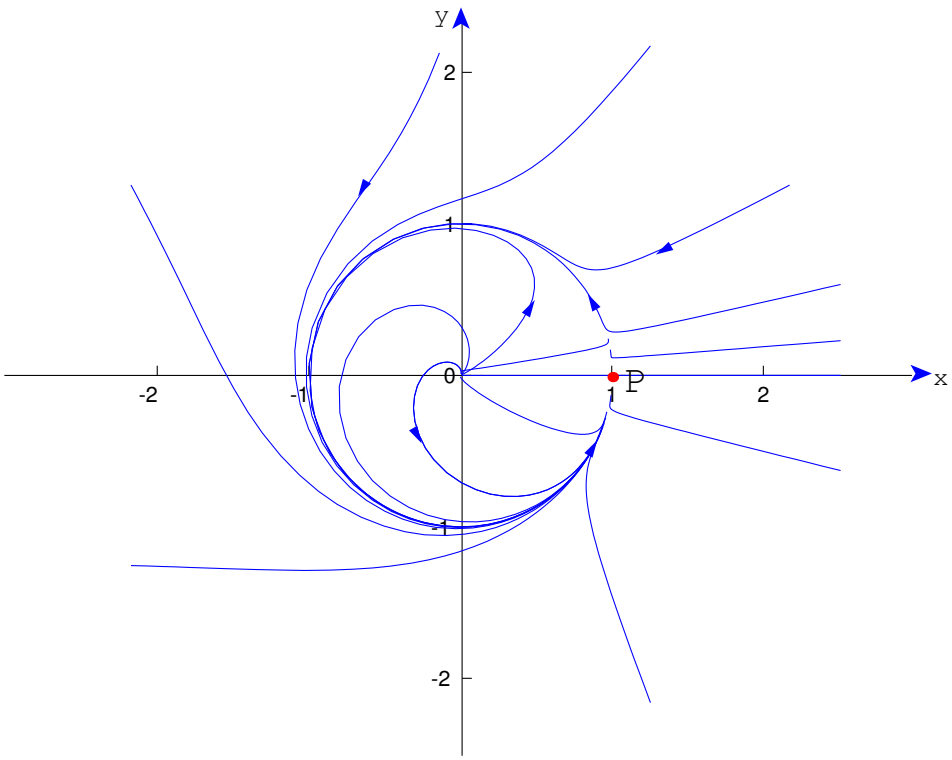

Figure 5: Attractivity without stability.

The steady state x^e is said to be *asymptotically stable* (We also say that (17) is asymptotically stable at x^e) if it is Lyapunov stable and attractive. It is *globally asymptotically stable* if it is Lyapunov stable and globally attractive.

Asymptotic stability of the equilibrium x^e means that all solutions starting near x^e not only stay nearby, but also tend to the equilibrium x^e as time goes to infinity.

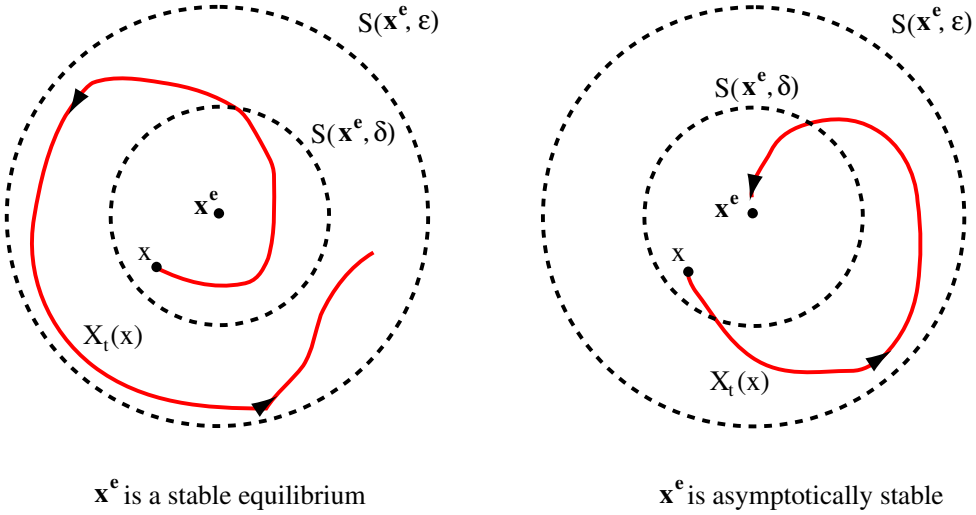

Figure 6: The phase portrait of a stable and an asymptotically stable equilibrium. $S(x^e, r)$ is a sphere centered at x^e with radius r.

The steady state x^e is said to be *marginally stable* if it is stable but not asymptotically stable. For instance, the origin is a marginally stable equilibrium point for the following planar system

$$\begin{cases} \dot{x} = -y \\ \dot{y} = x. \end{cases}$$

The system (17) is *exponentially stable* (respectively *globally exponentially stable*) at x^e if there exist two positive constants K an λ such that

$$\|X_t(x) - x^e\| < K\|x - x^e\|e^{-\lambda t}$$

for all x in a neighborhood of x^e (respectively for all $x \in \Omega$) and all positive time t.

An equilibrium x^e is an isolated equilibrium point for the system (17) if it has a neighborhood $\mathcal{U}$ such that $X(x) \neq 0$ for all $x \in \mathcal{U}$. In the sequel, only isolated equilibrium points will be considered and since any isolated equilibrium can be shifted to the origin (of $\mathbb{R}^n$) by a change of variables, $x' = x - x^e$, we shall assume that $x^e = 0$.

Remark: The above definitions mean that the solution of the differential system (17) has to be explicitly known before stability conclusions can be drawn. Except for a linear system, it is often not possible to compute the analytic expression of the solution of (17). To overcome this difficulty, some tools have been developed to explore the stability properties of nonlinear systems. The most effective are the linearization technique and the Lyapunov's second method. The linearization allows to derive only local conclusions while Lyapunov's second method allows to get global results.

3.2 Stability of linear systems

For linear time-invariant system

$$\begin{cases} \dot{x} = Ax, \\ x \in \mathbb{R}^n, \end{cases} \tag{18}$$

the solution is given by $X_t(x) = e^{tA}x$. Therefore, the previous stability conditions can be expressed in terms of the eigenvalues of the matrix A. Let $\sigma(A) = \{\lambda_i = \alpha_i + j\beta_i,\ i = 1 \dots k \leq n\}$ be the spectrum of A, i.e., the set of all the distinct eigenvalues of A, α_i is the real part of the eigenvalue λ_i and β_i the imaginary part. Let Ψ_A be the minimal polynomial of A, that is, the monic polynomial $\pi(\lambda)$ of least degree such that $\pi(A) = 0$. We have $\Psi_A(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \dots (\lambda - \lambda_k)^{n_k}$, with $n_1 + n_2 + \dots + n_k \leq n$. The stability properties of the system (18) can then be summarized as follows:

Theorem 3.1 *The system (18) is asymptotically stable if and only if $\alpha_i < 0$ for all eigenvalues. In this case, we say that A is a stable matrix or a Hurwitz matrix. If there is an eigenvalue λ_i for which $\alpha_i > 0$, then the system (18) is unstable. In this case the matrix A is called unstable.*

A matrix which is neither stable nor unstable is called *critical*. The eigenvalues of a critical matrix are all with non-positive real parts and at least one of them is with zero real part. The eigenvalues with vanishing real parts are called *critical eigenvalues* or *critical characteristic roots*.

Theorem 3.2 *If the matrix A is critical then the equilibrium of the equation (18) is marginally stable (stable but not attractive) if all the critical characteristic roots are simple roots of the minimal polynomial Ψ_A. Otherwise the equilibrium is unstable.*

For discrete-time system $x(k+1) = Ax(k)$, we have the following analogous criteria:

- the discrete-time system is unstable if there is an eigenvalue λ_i such that $|\alpha_i| > 1$ or if there is an eigenvalue λ_i which is not a simple root of the minimal polynomial Ψ_A ($n_i \geq 2$) for which $|\alpha_i| \geq 1$;
- the discrete-time system is Lyapunov stable if $|\alpha_i| \leq 1$ for all eigenvalues that are simple roots of Ψ_A and $|\alpha_i| < 1$ for all repeated eigenvalues;
- the discrete-time system is asymptotically stable if $|\alpha_i| < 1$ for all eigenvalues.

It must be noted that for linear systems (continuous or discrete-time), we have the following equivalences: asymptotic stability $\Longleftrightarrow$ global asymptotic stability $\Longleftrightarrow$ exponential stability.

3.3 Linearization and stability of nonlinear systems

Let $x = 0$ be an equilibrium point for the nonlinear system (17) where the vector function $X : \Omega \longrightarrow \mathbb{R}^n$ is assumed to be continuously differentiable and Ω is a neighborhood of the origin (i.e., the origin is an interior point of Ω). In a coordinate system $(x_1, x_2, \dots, x_n)$, the system (17) is given by the following system of differential equations:

$$\begin{cases} \dot{x}_1 = X_1(x) = X_1(x_1, x_2, \dots, x_n) \\ \vdots \\ \dot{x}_n = X_n(x) = X_n(x_1, x_2, \dots, x_n). \end{cases} \tag{19}$$

Let A be the Jacobian matrix of the vector function X at the origin, that is,

$$A = \frac{\partial X}{\partial x}(0) = \begin{pmatrix} \frac{\partial X_1}{\partial x_1}(0) & \frac{\partial X_1}{\partial x_2}(0) & \dots & \dots & \frac{\partial X_1}{\partial x_n}(0) \\ \frac{\partial X_2}{\partial x_1}(0) & \frac{\partial X_2}{\partial x_2}(0) & \dots & \dots & \frac{\partial X_2}{\partial x_n}(0) \\ \vdots & \vdots & & & \vdots \\ \frac{\partial X_n}{\partial x_1}(0) & \frac{\partial X_n}{\partial x_2}(0) & \dots & \dots & \frac{\partial X_n}{\partial x_n}(0) \end{pmatrix} \tag{20}$$

Then we can write

$\dot{x} = X(x) = Ax + Y(x)$

where the vector function Y satisfies $\lim_{\|x\|\to 0} \frac{\|Y(x)\|}{\|x\|} = 0$. The linear system $\dot{x} = Ax$ is called the *linearized* system of the nonlinear system (19) at the origin. This linear approximation is valid for small deviation from the equilibrium point $x = 0$. Often it is sufficient to draw conclusions about the local stability properties of the nonlinear system thanks to the following theorem known as *Lyapunov's first method*.

Theorem 3.3 *If all the eigenvalues of the Jacobian matrix A have negative real parts, then the origin is an asymptotically stable equilibrium point for the nonlinear system (19). If at least one eigenvalue of A has a positive real part, then the origin is unstable for the nonlinear system (19).*

The linearization technique or the Lyapunov's first method is a simple technique, and is usually the first method used in the stability analysis of an equilibrium point for a given nonlinear system. However, it does not say how large is the region of attraction of the considered equilibrium when all the eigenvalues of the Jacobian matrix have negative real parts. It also must be noted that the Lyapunov's first method is inconclusive when all the eigenvalues of the Jacobian matrix have non-positive real parts and some of them actually have a zero real part.

3.4 Lyapunov functions

The Lyapunov's first method uses the linearized system to reveal the stability properties of the nonlinear system (19). The *Lyapunov's second method* (or the *direct method*) works explicitly with the nonlinear system (19). This method can often be used to determine the stability of the equilibrium when the information obtained from the linearization is inconclusive. It also has the advantage of enabling the analysis to extend beyond only a small neighborhood of the equilibrium. The Second Method of Lyapunov is based on the use of auxiliary functions called *Lyapunov functions*.

Let V be a real-valued function defined and continuous in Ω. We say that the function V is a *Lyapunov function* for the system (17) on $\mathcal{U} \subset \Omega$ if it is non-increasing along the solutions of the system (17), that is, $V(X_t(x)) \leq V(X_{t'}(x))$ for all $x \in \mathcal{U}$ and all $0 \leq t \leq t'$. When the function V is of class C^1 then its value never increases along the trajectories of the system if the time derivative of $t \mapsto V(X_t(x))$ is non-positive, that is, for all $x \in \mathcal{U}$,

$\dot{V}(x) = X.V(x) \leq 0.$

The function $X.V$ is called the *Lie derivative* of V along the vector function X, it is defined by

$$X.V\ (x) = \frac{d}{dt}\Big(V\ \big(X_t(x)\big)\Big)\Big|_{t=0}.$$

Let $(x_1, \ldots, x_n)$ be a coordinate system. If $X(x) = (X_1(x), \ldots, X_n(x))^T$, $\langle .\ ,\ . \rangle$ is a scalar product and ∇V is the gradient of V in these coordinates, then

$$X.V\ (x) = \langle X(x)\ ,\ \nabla V(x) \rangle = \sum_{i=1}^{n} X_i(x) \frac{\partial V}{\partial x_i}(x).$$

This formula shows that the time derivative of the function V along the solutions of the system (17) can be computed without explicitly integrating the differential equation.

A real-valued function V is said to be *positive definite* at a point $x^e \in \Omega$ if there is a neighborhood $\mathcal{U}$ of x^e such that $V(x) > 0$ for each $x \in \mathcal{U} \setminus \{x^e\}$. Similarly, the function V is said to be *negative definite* at x^e on $\mathcal{U}$ if $V(x) < 0$ for all $x \in \mathcal{U} \setminus \{x^e\}$.

We go back to the system (17) and we suppose that the origin is an equilibrium point, i.e., $X(0) = 0$. The stability properties of the system (17) can be studied with the help of Lyapunov functions thanks to the following result known as *Lyapunov's second method*. This method can be seen as a generalization of the energy method. It has been inspired by the fact that a stable equilibrium state for a mechanical system corresponds to a local minimum of the total energy.

Theorem 3.4 (Lyapunov 1892) *Suppose there exists a C^1 function V defined on some neighborhood $\mathcal{U}$ of the origin such that*

i. *The function V is positive definite at the origin, i.e., $V(x) > 0$ for all $x \in \mathcal{U} \setminus \{0\}$ and $V(0) = 0$.*

ii. *$\dot{V}(x) \leq 0$ for all $x \in \mathcal{U}$.*

Then the origin is a Lyapunov stable equilibrium point for the system (17).
If moreover the function $x \mapsto \dot{V}(x)$ is negative definite, that is $\dot{V}(x) < 0$ for all $x \in \mathcal{U} \setminus \{0\}$ and that $\dot{V}(0) = 0$, then the origin is an asymptotically stable equilibrium point for the system (17).

There also exists a global version of this theorem. To this end we need to introduce the notion of a *proper* function. A continuous real valued function $V : \Omega \longrightarrow V(\Omega) \subset \mathbb{R}$ is said to be proper if, for all $\alpha \in V(\Omega) \subset \mathbb{R}$, the set $\{x \in \Omega : V(x) \leq \alpha\}$ is a compact subset of Ω. Proper functions are useful for proving that the solutions of the system (17) are bounded: *Suppose that the system (17) has a Lyapunov function V on Ω, that is $\dot{V}(x) \leq 0$, for all $x \in \Omega$. If the function V is proper then all the solutions are bounded, i.e.,*

$$\forall x \in \Omega,\ \exists M \geq 0 \quad \text{such that} \quad \|X_t(x)\| \leq M,\ \forall t \geq 0.$$

Since V is assumed to be continuous and Ω is a connected open subset of $\mathbb{R}^n$, we have $V(\Omega) = (a,\ b)$ (or $V(\Omega) = [a,\ b)$), b can be a real or $+\infty$. So, another characterization for the function V to be proper is that

$$\lim_{x \to \partial\Omega} V(x) = b, \quad \lim_{\substack{\|x\| \to +\infty \\ x \in \Omega}} V(x) = b$$

For instance, $V : x \mapsto \arctan(\|x\|)$ is proper on $\mathbb{R}^n$. The function $W : (x_1, x_2) \mapsto x_1^2 + \dfrac{x_2^2}{1+x_2^2}$ is not proper on $\mathbb{R}^2$ because $\lim\limits_{x_2 \to +\infty} W(0, x_2) = 1$.
Now, we give the global version of the Lyapunov's theorem:

Theorem 3.5 *If there exists a C^1 function V defined on the whole state space Ω and satisfying*
i. *The function V is positive definite at the origin, i.e, $V(x) > 0$ for all $x \in \Omega \setminus \{0\}$ and $V(0) = 0$.*
ii. *$\dot{V}(x) < 0$ for all $x \in \Omega \setminus \{0\}$.*
iii. *The function V is proper.*
Then the origin is a globally asymptotically stable equilibrium for the system (17).

Example: The Predator-Prey Model
Let us consider again the simple Lotka-Volterra predator-prey model (7) with $u = 0$:

$$\begin{cases} \dot{N} = aN - bNP \\ \dot{P} = -cP + ebNP \end{cases} \tag{21}$$

All the variables and constants are positive. This system has two equilibria: the trivial equilibrium $(0, 0)$ and the other equilibrium $(N^*, P^*) = (\dfrac{c}{eb}, \dfrac{a}{b})$.
The linearization of the system (21) around the point $(0, 0)$ leads to

$$\begin{pmatrix} \dot{N} \\ \dot{P} \end{pmatrix} = \underbrace{\begin{pmatrix} a & 0 \\ 0 & -c \end{pmatrix}}_{A_0} \begin{pmatrix} N \\ P \end{pmatrix}$$

The matrix A_0 has a positive eigenvalue which implies that the point $(0, 0)$ is an unstable equilibrium point.
The linearization of the system (21) around the point (N^*, P^*) leads to

$$\begin{pmatrix} \dot{N} \\ \dot{P} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & -c/e \\ ea & 0 \end{pmatrix}}_{A_1} \begin{pmatrix} N \\ P \end{pmatrix}.$$

The matrix A_1 has two eigenvalues whose real parts are zero, so the linearization method does not give any information about the stability of the nonlinear system (21).

In this situation, the Lyapunov's theorem can be helpful as we shall see. Let Ω be the interior of the positive quadrant, i.e., $\Omega = \mathbb{R}^2_{+>} = \{(N, P) \in \mathbb{R}^2 : N > 0, P > 0\}$. Let V be the function defined on Ω by

$$V(N, P) = e\,b\,N - c\,\log(\frac{e\,b}{c}N) + b\,P - a\,\log(\frac{b}{a}P) - (a + c).$$

This function is positive definite at (N^*, P^*): $V(N^*, P^*) = 0$ and $V(N, P) > 0$ for all $N \neq N^*$ and $P \neq P^*$. The function V is proper on Ω because we have

$$V(\Omega) = [0, +\infty), \lim_{\|(N,P)\| \to +\infty} V(N, P) = +\infty, \lim_{N \to 0} V(N, P) = +\infty \text{ and } \lim_{P \to 0} V(N, P) = +\infty.$$

Moreover its time derivative is $\dot{V} = 0$. Therefore, by Lyapunov's theorem, the nontrivial equilibrium (N^*, P^*) is stable and, since V is proper on Ω, all the solutions are bounded. Actually, the above function V, known as the *ecological Lyapunov function*, is a *constant* of motion. Hence, the solutions of (21) lie on fixed curves defined by $V = k$. For a given initial population distribution (N_0, P_0), the corresponding solution $(N(t), P(t))$ of (21) satisfies $V(N(t), P(t)) = V(N_0, P_0)$, for all positive time t.

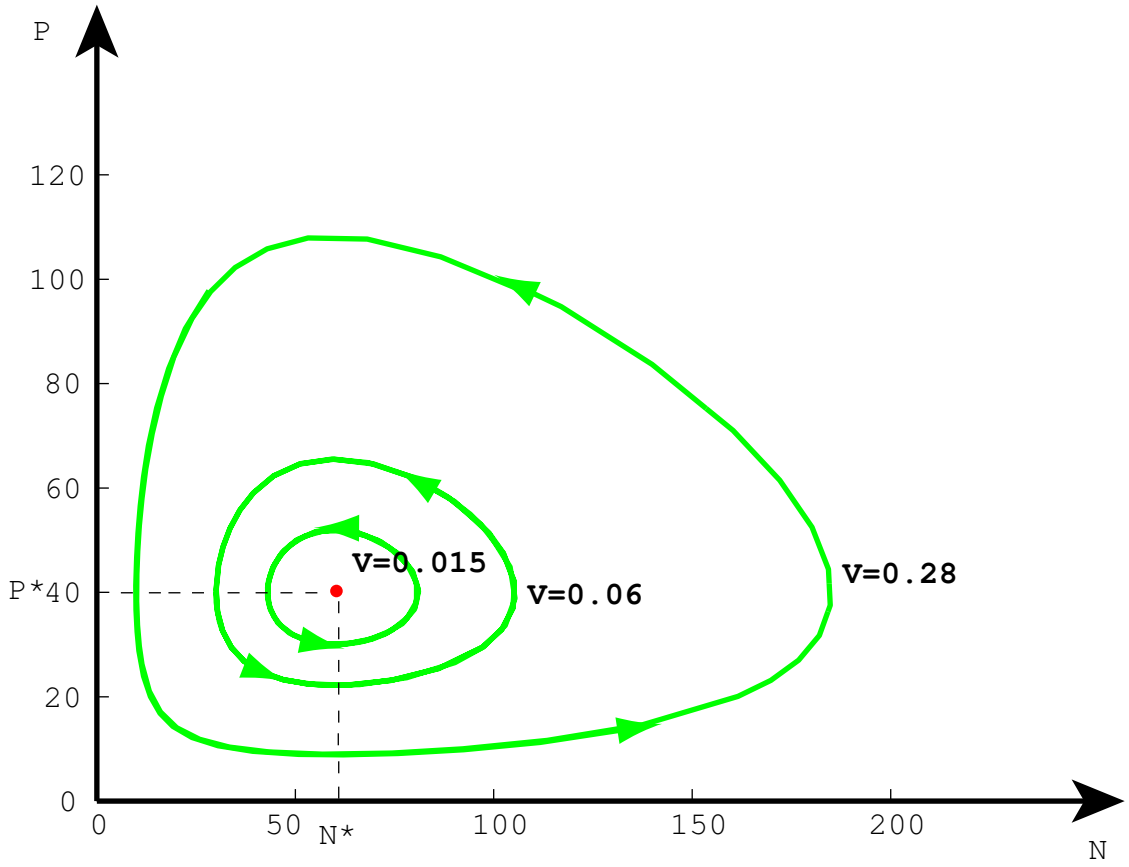

Figure 7: The trajectories of (21) for $a = 0.4$, $b = 0.01$, $c = 0.3$, $e = 0.5$.

Now, if we add to the previous predator-prey system terms representing decay due to crowding, we get

$$\begin{cases} \dot{N} = aN - bNP - fN^2 \\ \dot{P} = -cP + ebNP - gP^2 \end{cases} \tag{22}$$

The nontrivial equilibrium for this system is $N^* = \dfrac{ag + bc}{eb^2 + fg}$ and $P^* = \dfrac{abe - cf}{eb^2 + fg}$. This equilibrium belongs to Ω if and only if $abe - cf > 0$.
A candidate Lyapunov function is

$$V(N, P) = eN^* \left(\frac{N}{N^*} - \log\left(\frac{N}{N^*} \right) \right) + P^* \left(\frac{P}{P^*} - \log\left(\frac{P}{P^*} \right) \right) - eN^* - P^*.$$

This function is positive definite at (N^*, P^*) and it is proper on Ω. Let us evaluate its time

derivative along the solutions of the system (22):

$$\begin{aligned}
\dot{V}(N,P) &= e\left(1-\frac{N^*}{N}\right)\left(aN-bNP-fN^2\right)+\left(1-\frac{P^*}{P}\right)\left(-cP+ebNP-gP^2\right)\\
&= e\,(N-N^*)\left(a-bP-fN-\underbrace{(a-bP^*-fN^*)}_{=0}\right)\\
&+(P-P^*)\left(-c+ebN-gP-\underbrace{(-c+ebN^*-gP^*)}_{=0}\right)\\
&= e\,(N-N^*)\left(-f(N-N^*)-b(P-P^*)\right)+(P-P^*)(eb(N-N^*)-g(P-P^*))\\
&= -ef\,(N-N^*)^2-g(P-P^*)^2.
\end{aligned}$$

Hence $\dot{V}(N,P)<0$ for all $(N,P)\in\Omega\setminus\{(N^*,P^*)\}$ and $\dot{V}(N^*,P^*)=0$. Thus, by the global version of Lyapunov's theorem, we conclude that the nontrivial equilibrium (N^*,P^*) is globally asymptotically stable.

In order to use the original Lyapunov's theorems for the purpose of proving the asymptotic stability of a given system, we have to find a positive definite function V whose time-derivative $\dot{V}$ is negative definite. This is a difficult task in general. The definiteness of the derivative $\dot{V}$ can actually be relaxed by using the LaSalle's Invariance Principle that we expose hereafter.

A set $G\subset\Omega$ is an *invariant set* for the system (17) if whenever a solution belongs to G at some time, then it belongs to G for all future and past time. That is, if $x\in G$ then $X_t(x)\in G$, for all $t\in\mathbb{R}$. For example, the following sets are invariant sets for the predator-prey system (21):

- The positive quadrant,
- $\{(N,0), N\in\mathbb{R}\}$,
- $\{(0,P), P\in\mathbb{R}\}$,
- $\{(N,P) : e\,b\,N-c\,\log(\frac{e\,b}{c}N)+b\,P-a\,\log(\frac{b}{a}P)=k\}$, k being a positive constant.

A set G is *positively invariant* (*negatively invariant*) if:

$x\in G\Longrightarrow X_t(x)\in G,\ \forall t\geq 0\ (\forall t\leq 0)$.

The LaSalle's Invariance Principle can be stated as follows

Theorem 3.6 *Let Ω be a subset of $\mathbb{R}^n$. Assume that Ω is positively invariant for system (17). Let $V:\Omega\to\mathbb{R}$ be a C^1 scalar function such that $\dot{V}(x)\leq 0$ in Ω. Let E be the set of points within Ω where $\dot{V}(x)=0$, and let $\mathcal{L}$ be the largest invariant set within E. Then every bounded solution starting in Ω tends to the set $\mathcal{L}$ as time goes to infinity.*

This theorem is a very useful tool for system analysis. Unlike Lyapunov's theorem, it does require neither the function V to be positive definite nor the function $\dot{V}$ to be negative definite. However, it gives only information about the attraction. Therefore it can be used to show that the solutions tend to an equilibrium if the set $\mathcal{L}$ is reduced to this equilibrium but it does not allow to say if this equilibrium is stable or not. Actually, when we are interested in establishing asymptotic stability of an equilibrium point (assumed to be the origin of $\mathbb{R}^n$), we use the following corollary of the LaSalle's Invariance Principle:

Corollary 3.1 *Let Ω be an open connected subset of $\mathbb{R}^n$ such that $x = 0 \in \Omega$, and $x = 0$ is an equilibrium state for the system (17). Let $\mathcal{U}$ be a neighborhood of the origin in Ω and let $V : \mathcal{U} \to \mathbb{R}$ be a C^1 positive definite function, such that $\dot{V}(x) \leq 0$ in $\mathcal{U}$. Let $E = \{x \in \mathcal{U} : \dot{V}(x) = 0\}$, and assume that largest invariant set within E is reduced to the origin. Then, the origin is asymptotically stable.*
If the above conditions hold for $\mathcal{U} = \Omega$ and the function V is proper on Ω, then, the origin is a globally asymptotically stable equilibrium point for the system (17).

This result contains the original Lyapunov's theorem as a special case. It must be noted that the set Ω needs not to be bounded.

Example: Consider again the predator-prey system (21) and suppose that the crowding affects only the prey growth:

$$\begin{cases} \dot{N} = aN - bNP - fN^2, \\ \dot{P} = -cP + ebNP. \end{cases} \tag{23}$$

This system has two equilibria: the trivial equilibrium $N = P = 0$ (which is unstable) and a nontrivial equilibrium $N^* = \dfrac{c}{eb}$ and $P^* = \dfrac{abe - cf}{eb^2}$. This equilibrium belongs to Ω (The interior of the positive quadrant) if and only if $abe - cf > 0$. We take again the following candidate Lyapunov function

$$V(N,P) = eN^* \left(\frac{N}{N^*} - \log\left(\frac{N}{N^*}\right)\right) + P^* \left(\frac{P}{P^*} - \log\left(\frac{P}{P^*}\right)\right) - eN^* - P^*.$$

Its time-derivative along the solutions of (23) is

$$\dot{V}(N,P) = -ef\,(N - N^*)^2 \leq 0.$$

In this case $\dot{V}$ is not negative definite so Lyapunov's theorem does not apply. Hence, we have to apply the LaSalle's Invariance principle. Here, $E = \{(N,P) \in \Omega : \dot{V}(N,P) = 0\} = \{(N^*, P), P > 0\}$. On E, the vector field is $\begin{pmatrix} \dot{N} \\ \dot{P} \end{pmatrix} = \begin{pmatrix} (a - bP - fN^*)N^* \\ 0 \end{pmatrix}$ (cf Figure 8). It is then easy to see that the only invariant set within E is the equilibrium point (N^*, P^*) and this proves that this equilibrium is globally asymptotically stable.

When the vector field X and the Lyapunov function are analytic, the set $\mathcal{L}$ can be computed according to the following formula:

$$\mathcal{L} = \{x \in \Omega : X^k.V(x) = 0,\ k = 1, 2, \ldots\}.$$

Hence, we just have to solve a system of equations. For the previous system (23), we have, with $x = (N,P)$ and $X = (aN - bNP - fN^2, -cP + ebNP)^T$,

$$\begin{array}{ll} X.V(x) = 0 \Longrightarrow N - N^* = 0 & (i) \\ X^2.V(x) = -2ef(N - N^*)(aN - bNP - fN^2) = 0 & (ii) \\ X^3.V(x) = -2ef(aN - bNP - fN^2) - 2ef(N - N^*)(-bN)(-cP + ebNP) & \\ +(a - bP - 2fN)(-2ef(N - N^*)(aN - bNP - fN^2) = 0 & (iii) \end{array}$$

By combining (i) and (iii), we get $N = N^*$ and $P = P^*$.

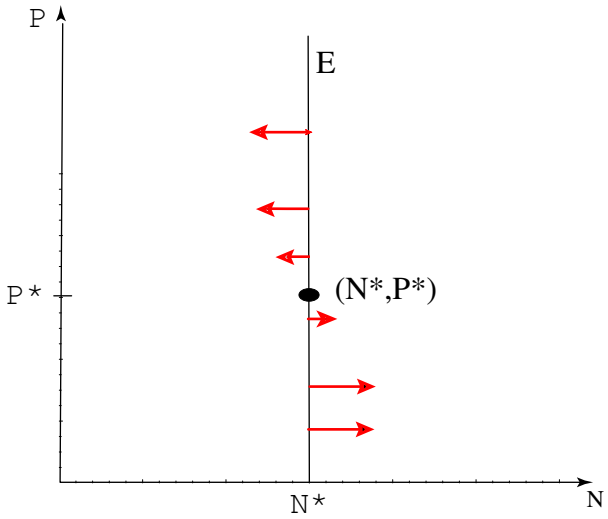

Figure 8: The set E where $\dot{V} = 0$.

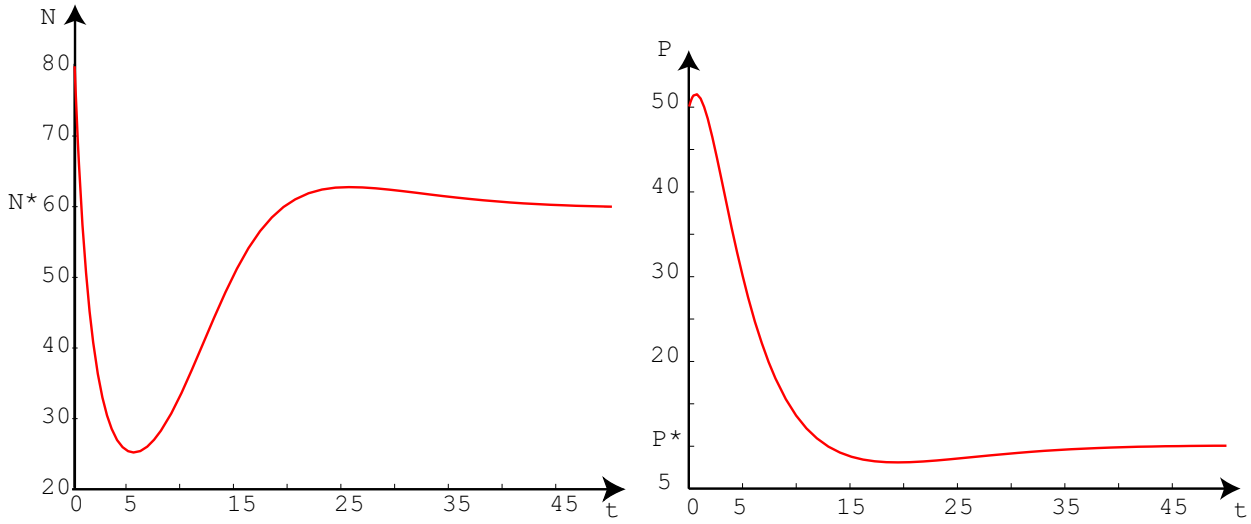

Figure 9: The time evolution of the prey and the predator governed by (23).

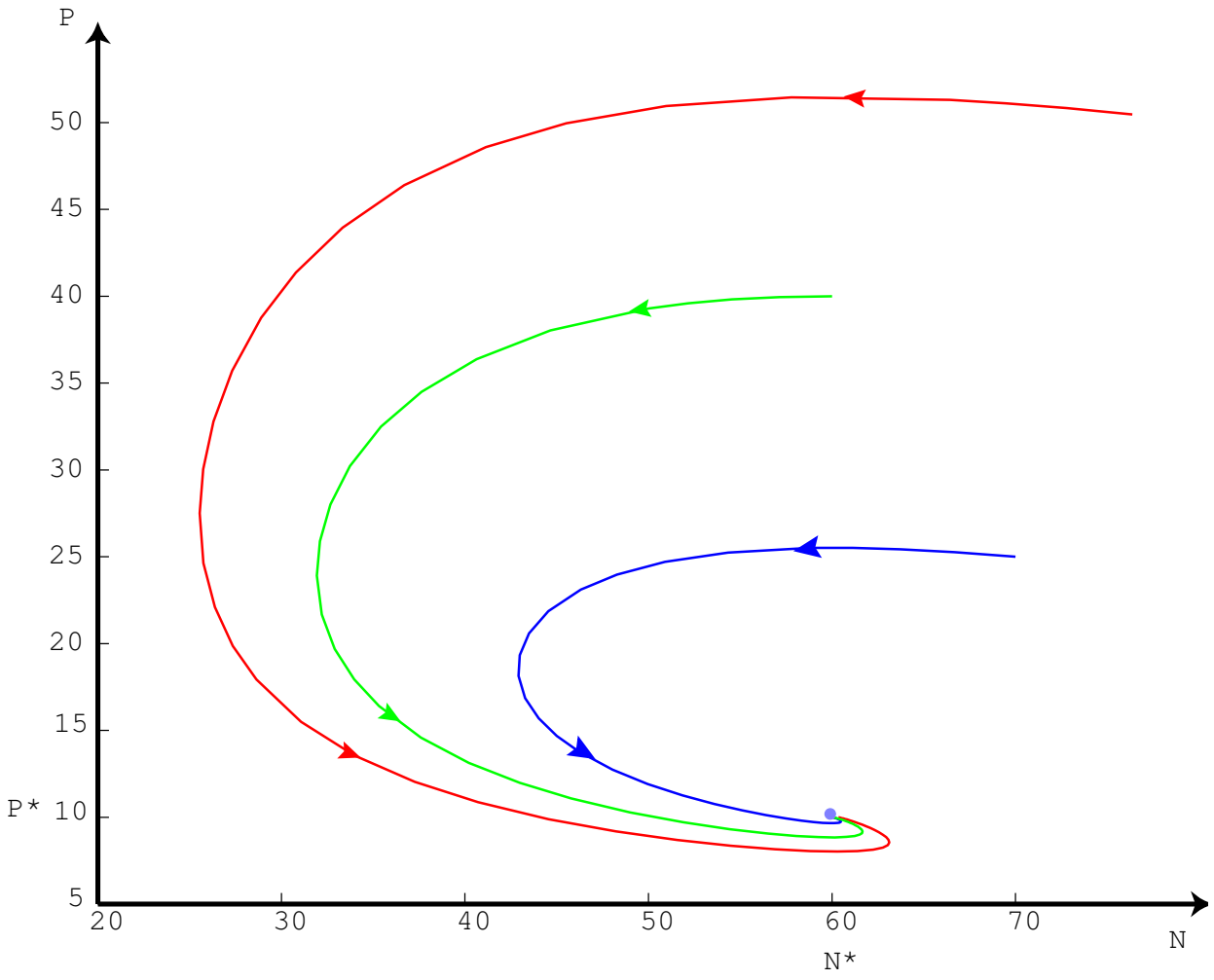

Figure 10: The trajectories of (23) corresponding to different initial conditions.

3.5 Limit cycle

Thus far the only type of attractors encountered have been equilibrium points. For nonlinear systems, an other type of attractor is possible, the *limit cycle*.

Here we consider two-dimensional systems of the form

$$\begin{cases} \dot{x}_1 = X_1(x_1, x_2), \\ \dot{x}_2 = X_2(x_1, x_2), \\ x = (x_1, x_2) \in \Omega \subset \mathbb{R}^2, \quad X(x) = (X_1(x), X_2(x)), \end{cases} \tag{24}$$

where X_1 and X_2 are continuously differentiable.

A solution of (24) through the point $x \in \Omega$ is said to be *periodic* if there exists $T > 0$ such that

$$X_{t+T}(x) = X_t(x), \quad \forall t \in \mathbb{R}. \tag{25}$$

The trajectory corresponding to a periodic solution is called a *periodic orbit* or a *closed orbit*. The period of a periodic solution is the smallest $T > 0$ such that (25) holds. A constant solution

(equilibrium position) is a trivial periodic solution. Here we are interested in nontrivial periodic solutions. A system *oscillates* when it has a (nontrivial) periodic solution.

Consider the linear two-dimensional system $\dot{x}_1 = -x_2,\ \dot{x}_2 = x_1$. All the solutions are periodic and the corresponding orbits are circles centered at the origin. One can remark that there is a continuum of periodic orbits and that the amplitude of the oscillations depend on the initial condition. The above system is an *harmonic oscillator*.

Now consider the following nonlinear two-dimensional system

$$\begin{cases} \dot{x}_1 = -x_2 + x_1(1 - x_1^2 - x_2^2), \\ \dot{x}_2 = x_1 + x_2(1 - x_1^2 - x_2^2). \end{cases} \tag{26}$$

The origin is an equilibrium point and there is no other equilibrium point. The linearization of the system (26) at the origin is given by the matrix

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

This matrix has two eigenvalues whose real parts are positive. Therefore the origin is an unstable equilibrium point for (26).

The unit circle is an invariant set of the system (26). This fact can be easily seen by representing the system in the polar coordinates (r, θ) defined by:

$x_1 = r\cos\theta,\ x_2 = r\sin\theta,$

which yields

$$\begin{cases} \dot{r} = r(1 - r^2), \\ \dot{\theta} = 1. \end{cases}$$

This shows that the sets defined respectively by $r = 0$ and $r = 1$ are invariant sets of the system. The first set $r = 0$ is just the origin and the second $r = 1$ corresponds to the unit circle. The polar form also shows that the unit circle is the unique periodic orbit for the system (26). One also can remark that if $r < 1$ then $\dot{r} > 0$ and if $r > 1$ then $\dot{r} < 0$. Therefore all the trajectories (except the trivial solution $X_t(0) \equiv 0$) spiral toward the unit circle from inside or outside (see Figure 11). The unit circle is a *stable limit cycle*.

Remark: It is possible to show that the unit circle attract all the solutions (except the trivial one) of the system (26) by considering the real-valued function $V(x_1, x_2) = (1 - x_1^2 - x_2^2)^2$ and applying the LaSalle's Invariance Principle (Theorem 3.6).

A closed orbit γ is called a *limit cycle* if it is an isolated closed orbit, that is, there exists a neighborhood of γ which contains no other closed orbits of the system (24). A limit cycle is a special periodic solution of the planar system (24) that attract all other nearby solutions as $t \to \infty$ or $t \to -\infty$. Geometrically this means that the nearby non-closed trajectories spiral toward it, either from inside or outside as $t \to \infty$ or as $t \to -\infty$.

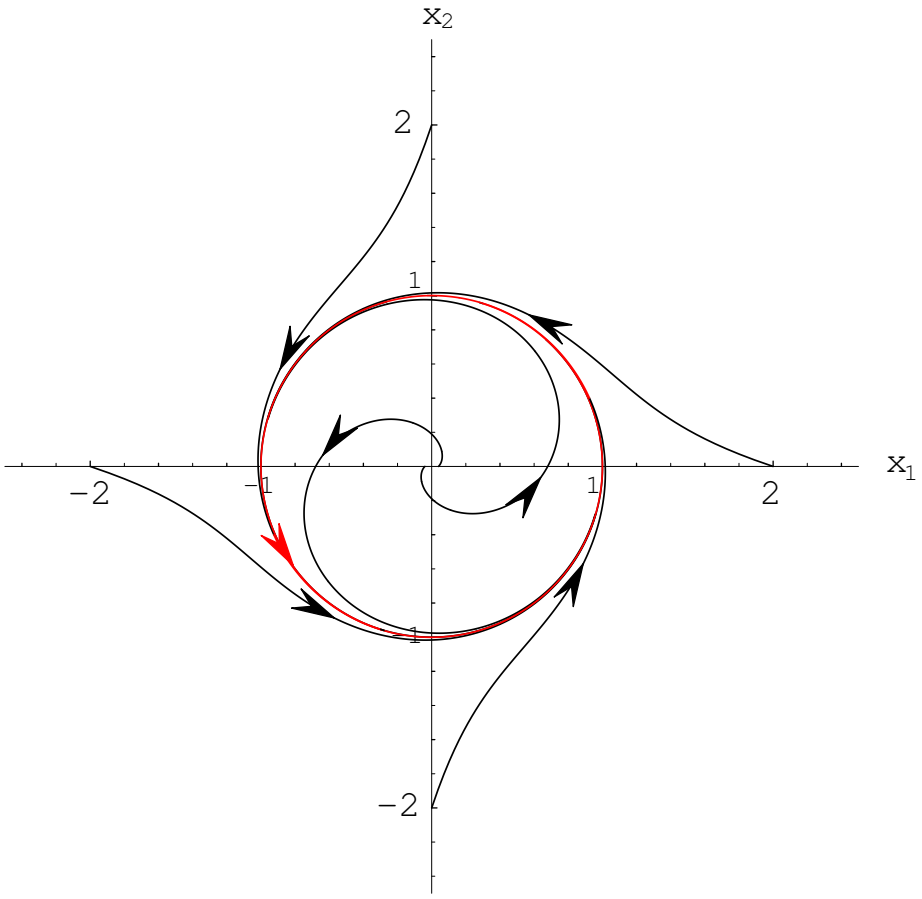

Figure 11: The trajectories of (26) corresponding to different initial conditions: the unit circle is a limit cycle.

A limit cycle is called stable when nearby trajectories approach it as time t approaches infinity. It is unstable if the all the trajectories starting arbitrarily close to it will tend away from it as $t \to \infty$.

The unit circle is a stable limit cycle for the system (26). It is an unstable limit cycle for

$$\begin{cases} \dot{x}_1 = x_2 - x_1(1 - x_1^2 - x_2^2), \\ \dot{x}_2 = -x_1 - x_2(1 - x_1^2 - x_2^2). \end{cases} \tag{27}$$

This system is the same as (26) in reversed time. Therefore, its phase portrait is the same as the one of (26) but the arrow heads are reversed.

We present a celebrated result known as Poincaré-Bendixson theorem which is very useful for establishing the existence of periodic orbits, and for exploring the stability properties of nonlinear two dimensional systems. It must be noted however that it has no generalization to higher dimensional systems.

Theorem 3.7 (Poincaré-Bendixson) *Let $\mathcal{D}$ be a positively invariant compact set for the planar system (24) containing a finite number of equilibria. Let x be a point of $\mathcal{D}$, and consider the corresponding solution $X_t(x)$ of (24). Then one of the following possibilities holds.*

i. *The solution $X_t(x)$ is a periodic solution;*
ii. *the solution $X_t(x)$ approaches a periodic solution as $t \to \infty$;*
iii. *the solution $X_t(x)$ approaches a cycle graph as $t \to \infty$.*

A cycle graph is the union of equilibria and trajectories $X_t(y)$ such that $\lim_{t\to-\infty} X_t(y)$ and $\lim_{t\to\infty} X_t(y)$ are equilibria. A cycle graph can be seen as the union of equilibria and orbits connecting them.

The Poincaré-Bendixson theorem says that a bounded trajectory that does not approach any equilibrium point is either a closed periodic orbit or approaches a closed periodic orbit as $t \to \infty$.

An implication of this theorem is that *a nonempty compact set that is positively or negatively invariant contains either an equilibrium point or a limit cycle.*

The following criterion is practical in ruling out the presence of closed orbits in a region of the plane.

Theorem 3.8 (Bendixson's negative criterion) *If on a simply connected domain $\mathcal{D} \subset \Omega$ (that is $\mathcal{D}$ is a connected set without holes) the expression $\dfrac{\partial X_1}{\partial x_1} + \dfrac{\partial X_2}{\partial x_2}$ is not identically zero and does not change sign, then the system (24) has no closed orbits lying entirely in $\mathcal{D}$.*

A consequence of the above criterion is that limit cycles can only be obtained with nonlinear systems.

Another criterion that can be used to preclude the existence of periodic solutions in a region of the plane when the Bendixsons fails is the following result due to Dulac.

Theorem 3.9 (Dulac's negative criterion) *Suppose there exists a real-valued function ρ, continuously differentiable in a simply connected domain $\mathcal{D} \subset \Omega$, such that the expression $\dfrac{\partial(\rho X_1)}{\partial x_1} + \dfrac{\partial(\rho X_2)}{\partial x_2}$ is not identically zero and does not change sign, then the system (24) has no closed orbits lying entirely in $\mathcal{D}$.*

Example: Consider the following simple epidemic model which is a simplified version of the system (2) introduced in Section 1 without the class E and without a disease-related death ($d = 0$).

$$\begin{cases} \dot{S} = bN - \mu S - \beta \dfrac{SI}{N}, \\ \dot{I} = \beta \dfrac{SI}{N} - (r+\mu)I, \\ \dot{T} = rI - \mu T, \\ \dot{N} = (b-\mu)N. \end{cases} \tag{28}$$

The proportions $s = S/N$, $i = I/N$, and $\tau = T/N$ satisfy the following system of differential equations

$$\begin{cases} \dot{s} = b - bs - \beta si, \\ \dot{i} = \beta si - (r+b)i, \\ \dot{\tau} = ri - b\tau. \end{cases} \tag{29}$$

Since $s + i + \tau = 1$ it is sufficient to study the planar system

$$\begin{cases} \dot{s} = b - bs - \beta si = X_1(s,i), \\ \dot{i} = \beta si - (r+b)i = X_2(s,i). \end{cases} \tag{30}$$

The system evolves in the set $\mathcal{D} = \{(s,i) \in \mathbb{R}^2 \ : 0 \leq s \leq 1,\ 0 \leq i \leq 1,\ 0 \leq s+i \leq 1\}$ and has two equilibria:

The disease-free equilibrium: $s^* = 1$, $i^* = 0$.

The endemic equilibrium: $s^* = \dfrac{b+r}{\beta} = \dfrac{1}{R_0}$, $i^* = \dfrac{b\ (\beta - (b+r))}{(b+r)\ \beta}$. This equilibrium belongs to $\mathcal{D}$ only if $b + r < \beta$.

The linearization technique allows to get the following local stability properties: The disease-free equilibrium ($s^* = 1$, $i^* = 0$) is asymptotically stable if $b + r > \beta$ and it is unstable if $b+r < \beta$. The endemic equilibrium is asymptotically stable when it exists, that is, if $b+r < \beta$. To rule out the existence of periodic orbits we first apply the Bendixson's criterion on $\mathcal{D}$ which is simply connected:

$$\frac{\partial X_1}{\partial s} + \frac{\partial X_2}{\partial i} = -b - \beta i + \beta s - (r+b) \leq -b - \beta i + \beta - (r+b) \text{ since } s \leq 1.$$

Therefore the planar system (30) has no periodic solutions if $b + r > \beta$. When $b + r \leq \beta$, Bendixson's criterion does not allow to conclude. We apply the Dulac's criterion with the function $\rho(s,i) = \frac{1}{si}$ which is continuously differentiable in $\mathcal{D}_1 = \{(s,i) \in \mathcal{D} : s > 0,\ i > 0\}$:

$$\frac{\partial(\rho X_1)}{\partial s} + \frac{\partial(\rho X_2)}{\partial i} = -\frac{b}{s^2 i} < 0,\ \forall (s,i) \in \mathcal{D}_1.$$

Thus, for any values of the positive parameters, the system has no periodic orbits lying entirely in $\mathcal{D}_1$. The set $\mathcal{D} \setminus \mathcal{D}_1$ can not contain a periodic orbit since it is just the union of two segments. Therefore there are no periodic orbits in $\mathcal{D}$.

Now the Poincaré-Bendixson theorem allows to get the following global stability properties:

- For any values of the parameters, the set $\mathcal{D}_0 = \{(s,i) \in \mathcal{D} :\ i = 0\}$ is positively invariant and the disease-free equilibrium attracts all the trajectories emanating from this set.
- If $b + r \geq \beta$ then the disease-free equilibrium ($s^* = 1$, $i^* = 0$) is globally attractive in $\mathcal{D}$: the disease dies out.
- If $b + r < \beta$ then the endemic equilibrium is globally asymptotically stable in the region $\mathcal{D} \setminus \mathcal{D}_0 = \{(s,i) \in \mathcal{D} :\ i > 0\}$: the disease will spread.

3.6 Stabilization

Now, we consider a controlled system with an equilibrium state at the origin of $\mathbb{R}^n$

$$\begin{cases} \dot{x}(t) = X(x(t), u(t)), \\ X(0,0) = 0, \\ x(t) \in \mathbb{R}^n, \ u(t) \in U \subset \mathbb{R}^m. \end{cases} \tag{31}$$

We assume that the function X is sufficiently smooth and we want to know if it is possible to control this system in order to stabilize it at its equilibrium state. There are two ways to achieve this goal. The first one consists in finding the control u as a function of the time variable t and corresponds to what is called *an open-loop control*. The second strategy is to build the stabilizing control as a function $u(x)$ of the state x, this is a *feedback control* or a *closed-loop control*. Here, we develop the basis of the second strategy.

We shall say that the system (31) is *stabilizable* if there exists a state feedback control $u(x)$, which is at least continuous (as a function of the state x), and such that the origin is an asymptotically stable equilibrium point for the closed-loop system

$$\dot{x}(t) = X(x(t), u(x(t))).$$

The function $u(x)$ is called a *static stabilizing feedback law*. The system (31) can also be *dynamically stabilizable* by use of a *dynamic feedback*. This means that we control the system by using other systems. So, we add to the system (31) another dynamical system whose input is x. Hence, we obtain an interconnected system:

$$\begin{cases} \dot{x} &= X(x,u) \\ \dot{y} &= G(y,x), \quad y \in R^q \end{cases} \tag{32}$$

We shall then say that the system (31) is dynamically stabilizable if there exists a feedback law $u(x,y)$ that stabilizes the system (32) at $(x,y) = (0,0)$. In this article, emphasis will be put on the static stabilization.

A necessary condition for the existence of a C^1 static feedback $u(x)$ that stabilizes locally the system (31) around its equilibrium point 0 is that

(i) there exists a neighborhood $\mathcal{N}$ of the origin such that for each state $\xi \in \mathcal{N}$ there is a control u_ξ steering the system from $x = \xi$ at $t = 0$ to $x = 0$ at $t = \infty$, i.e., $\lim\limits_{t \to +\infty} X_t^{u_\xi}(\xi) = 0$;

(ii) the image of the map

$$\begin{array}{rccc} X: & \mathbb{R}^n \times U & \longrightarrow & \mathbb{R}^n \\ & (x,u) & \longmapsto & X(x,u) \end{array}$$

contains a neighborhood of the origin;

(iii) the linearized system has no uncontrollable modes (see Controllability of linear systems 2.2) associated with eigenvalues whose real part is positive.

A system $\dot{x} = X(x,u)$ satisfying the condition **(i)** is said to be *asymptotically controllable to the origin* or open-loop stabilizable.

The conditions **(i)** and **(ii)** are necessary even if one requires that the stabilizer $u(x)$ is only continuous. The condition **(iii)** is not necessary if the stabilizer is required to be only continuous. When the condition **(i)** holds but the condition **(ii)** is not satisfied, then it is not

possible to stabilize the system (31) by a continuous feedback law $u(x)$; however, it is possible to stabilize it by a time-varying feedback $u(x,t)$ if some technical conditions hold. Moreover, for a given positive number $T > 0$, it is possible to choose a periodic stabilizer $u(x,t)$ that satisfies: $u(x, t+T) = u(x)$, for all $t \geq 0$.

Examples: We present some examples to illustrate the above obstructions to stabilizability:

1. The first example (due to R. Brockett) concerns a system that satisfies the conditions **(i)** and **(iii)** but does not satisfy the second condition **(ii)**:

$$\dot{x} = X(x,u), \text{ with } (x_1,x_2,x_3) \in \mathbb{R}^3, \ (u_1,u_2) \in \mathbb{R}^2, \quad \text{and } X(x,u) = \begin{pmatrix} u_1 \\ u_2 \\ x_2u_1 - x_1u_2 \end{pmatrix} \tag{33}$$

On the one hand, this driftless three-dimensional system is controllable (see Theorem 2.4). Hence the condition **(i)** is satisfied. On the other hand, the condition **(iii)** is obviously fulfilled since the linearized system at the origin

$$\begin{cases} \dot{x}_1 = u_1 \ , \\ \dot{x}_2 = u_2 \ , \\ \dot{x}_3 = 0 \ , \end{cases}$$

has two controllable modes and one uncontrollable mode associated with a vanishing eigenvalue. However the condition **(ii)** is not satisfied since it is not possible to find $(x,u) \in \mathbb{R}^3 \times \mathbb{R}^2$ such that $X(x,u) = (0,0,\epsilon)$ with $\epsilon \neq 0$. Therefore, the system (33) can not be stabilized at the origin by means of a continuous feedback law $u(x)$.

2. Consider the system

$$\begin{cases} \dot{x}_1 = x_1(x_1^2 + x_2^2) \ , \\ \dot{x}_2 = u \ , \\ (x_1,x_2) \in \mathbb{R}^2, \ u \in \mathbb{R} \ . \end{cases} \tag{34}$$

The conditions **(ii)** and **(iii)** are not violated but the condition **(i)** is not satisfied because for any $\alpha > 0$, the origin can not be asymptotically reached from any point of the open half space $x_1 > \alpha$. Hence there exists no continuous feedback law $u(x_1,x_2)$ which makes the origin asymptotically stable for the system (34).

3. The two-dimensional system

$$\begin{cases} \dot{x}_1 = x_1 - x_2^3 \ , \\ \dot{x}_2 = u \ , \\ (x_1,x_2) \in \mathbb{R}^2, \ u \in \mathbb{R} \ , \end{cases} \tag{35}$$

satisfies the conditions **(i)** and **(ii)** but does not satisfy the condition **(iii)** because the linearized system at the origin

$$\begin{cases} \dot{x}_1 = x_1 \ , \\ \dot{x}_2 = u \ , \end{cases}$$

has an uncontrollable mode associated with a positive eigenvalue. Therefore, the system (35) can not be stabilized at the origin by means of a C^1 feedback law $u(x_1,x_2)$.

3.6.1 Sufficient stabilizability conditions

For linear systems, the stabilization problem has been completely solved and the construction of static feedback stabilizers can be done in a systematic way. For nonlinear systems, the problem is a hard task and still under intensive investigation. No general method is available but many powerful techniques have been developed, each one allows to compute the stabilizers for a class of nonlinear systems. Here, we shall expose just one of these techniques because it is one of the most general and very simple to apply.
We consider an affine control system defined by smooth (C^∞) vector fields X, $Y_1, \ldots, Y_m$, with an equilibrium point at the origin of $\mathbb{R}^n$:

$$\begin{cases} \dot{x} = X(x) + \sum_{i=1}^{m} u_i Y_i(x). \\ X(0) = 0,\ Y_i(0) = 0. \end{cases} \tag{36}$$

If there exists a positive definite and proper smooth function $V : \mathbb{R}^n \to \mathbb{R}$ such that:
(i) The Lie-derivative of V with respect to the vector field X satisfies

$$X \cdot V(x) \leq 0,\ \forall x \in \mathbb{R}^n,$$

(ii) The set $W = \{x \in \mathbb{R}^n \mid X^{k+1} \cdot V(x) = X^k \cdot Y_i \cdot V(x) = 0,\ k \in \mathbb{N},\ i = 1, \ldots, m\}$ is reduced to the set $\{0\}$.
Then the affine system (36) is globally stabilizable by the smooth state feedback control law

$$u(x) = -\begin{pmatrix} Y_1 \cdot V(x) \\ \vdots \\ Y_m \cdot V(x) \end{pmatrix}. \tag{37}$$

Remark: The above feedback is known as *the Jurdjevic-Quinn feedback*. It must be noted that if the assumptions (i) and (ii) are satisfied then, for any smooth real-valued function β satisfying $\beta(x) > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$, the feedback law

$$u(x) = -\beta(x) \begin{pmatrix} Y_1 \cdot V(x) \\ \vdots \\ Y_m \cdot V(x) \end{pmatrix} \tag{38}$$

also stabilizes the system (36) at the origin because the time derivative of the function V along the solutions of the closed-loop system (36-38) is

$$\dot{V}(x) = X \cdot V(x) - \beta(x) \sum_{i=1}^{m} (Y_i \cdot V(x))^2 \leq 0$$

and the assumption (ii) implies that the largest invariant set contained in the set $\{\dot{V} = 0\}$ is reduced to the origin. Hence the LaSalle's Invariance Principle allows to conclude that the origin is a globally asymptotically stable equilibrium for the closed loop-system (36-38).

Example: A harvested fish population. Consider the following stage-structured model of a fish population that has been built by Touzeau-Gouzé. It includes $(n+1)$ stages represented by their abundance $x_i(t)$, stage 0 being the pre-recruits stage. Each stage is characterized by its fecundity, mortality and predation rates:

$$\begin{cases} \dot{x}_0(t) &= -\alpha x_0(t) - m_0 x_0(t) + \sum_{i=1}^{n} f_i l_i x_i(t) - \sum_{i=0}^{n} p_i x_i(t) x_0(t), \\ \dot{x}_1(t) &= \alpha x_0(t) - \alpha x_1(t) - m_1 x_1(t), \\ \quad \vdots & \\ \dot{x}_n(t) &= \alpha x_{n-1}(t) - \alpha x_n(t) - m_n x_n(t), \end{cases} \tag{39}$$

where

- m_i: linear mortality rate.
- α: linear aging coefficient.
- p_0: juvenile competition parameter.
- p_i: predation rate of class i on class 0.
- f_i: fecundity rate of class i.
- l_i: reproduction efficiency of class i.

The mortality coefficient can be written as a sum of the natural mortality rate M_i and the fishing mortality coefficient F_i. Hence one can write: $m_i = M_i + q_i E$, where q_i is the catchability of stage i and E is the fishing effort that can be seen as a control term. Denoting $\alpha_i = \alpha + M_i$, we get

$$\begin{cases} \dot{x}_0(t) &= -\alpha_0 x_0(t) + \sum_{i=1}^{n} f_i l_i x_i(t) - \sum_{i=0}^{n} p_i x_i(t) x_0(t) \\ \dot{x}_1(t) &= \alpha x_0(t) - (\alpha_1 + q_1 E(t)) x_1(t) \\ \quad \vdots & \\ \dot{x}_n(t) &= \alpha x_{n-1}(t) - (\alpha_n + q_n E(t)) x_n(t) \end{cases} \tag{40}$$

The origin is an equilibrium point which corresponds to an extinct population and is therefore not very interesting. Under some conditions and for a constant fishing effort $\bar{E}$, there exists another nontrivial equilibrium x^* whose coordinates are

$$\begin{cases} x_0^* = \dfrac{\sum_{i=1}^{n} f_i l_i \pi_i - \alpha_0}{p_0 + \sum_{i=1}^{n} p_i \pi_i} \\ x_i^* = \pi_i x_0^* \\ \pi_i = \dfrac{\alpha^i}{\prod_{j=1}^{i} (\alpha_i + q_i \bar{E})} \end{cases} \tag{41}$$

The goal is to compute the fishing effort as feedback control $E(x) = \bar{E} + u(x)$, in order to maintain the fish population around its steady state x^*. The state x^* becomes a globally asymptotically stable equilibrium for the closed-loop system within Ω, the positive quadrant. The system can be rewritten as follows

$$\dot{x} = \underbrace{\begin{pmatrix} -\alpha_0 x_0 + \sum_{i=1}^{n} f_i l_i x_i - \sum_{i=0}^{n} p_i x_i x_0 \\ \alpha x_0 - (\alpha_1 + q_1 \bar{E}) x_1 \\ \vdots \\ \alpha x_{n-1} - (\alpha_n + q_n \bar{E}) x_n \end{pmatrix}}_{X(x)} + u \underbrace{\begin{pmatrix} 0 \\ -q_1 x_1 \\ \vdots \\ -q_n x_n \end{pmatrix}}_{Y(x)}. \tag{42}$$

Let V be the following candidate Lyapunov function

$$V(x) = \frac{1}{2}\left((x_0 - x_0^*)^2 + \sum_{i=1}^{n}\left(\frac{\sum_{j=i}^{n} k_j\pi_j}{\alpha_i + q_i\bar{E}}\right)\left(\frac{x_i - x_i^*}{\pi_i}\right)^2\right), \quad k_i = f_i l_i - p_i x_0^* \qquad i = 1, \ldots, n.$$

V is a positive definite function on Ω provided that

$$x_0^* < \min_{i=1,\ldots,n} \frac{\sum_{j=i}^{n} f_j l_j \pi_j}{\sum_{j=i}^{n} p_j \pi_j}$$

Its derivative along the drift vector field X satisfies

$$X.V(x) = \langle X(x), \nabla V(x)\rangle \leq -\frac{1}{2}\sum_{i=1}^{n} k_i\pi_i\left((x_0 - x_0^*) - (\frac{x_i - x_i^*}{\pi_i})\right)^2 \leq 0$$

A candidate stabilizer given by (37) is $u = \Phi(x) = -\langle Y(x), \nabla V(x)\rangle = \sum_{i=1}^{n} \gamma_i q_i x_i (x_i - x_i^*)$, with $\gamma_i = \frac{\sum_{j=i}^{n} k_j\pi_j}{(\alpha_i + q_i\bar{E})\pi_i^2}$. However, the function Φ is unbounded and takes positive as well as negative values, hence $E(x) = \bar{E} + u(x)$ can take negative values which is not possible in practice (E is a fishing effort). Thus, instead of using the feedback given by the formula (37), we use the feedback law given by the formula (38) $u(x) = \beta(x)\Phi(x)$ and we choose the function β such that $\beta(x)\Phi(x) \geq -\bar{E}$. A simple computation shows that $\Phi(x) \geq -\sum_{i=1}^{n} \gamma_i q_i \frac{x_i^*}{4}$. Therefore we take $\beta(x) = \dfrac{4\bar{E}}{\sum_{i=1}^{n} \gamma_i q_i x_i^*}$ and

$$u(x) = \frac{4\bar{E}}{\sum_{i=1}^{n} \gamma_i q_i x_i^*}\Phi(x)\ . \tag{43}$$

We then have $u(x) \geq -\bar{E}$ which ensures that $E(x) \geq 0$ for all $x \in \Omega$. The figure 12 presents a simulation for a five stages system controlled by the feedback law given by the formula (43).

Stage i	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
p_i	0.2	0	0.1	0.1	0.1
f_i		0	0.5	0.5	0.5
l_i		0	10	20	15
q_i	0	0	0	0.1	0.15
α			0.8		
α_i	1.3	1	1	0.9	0.85
$\bar{E}$			1		

Table 1: The parameter values used in the simulation.

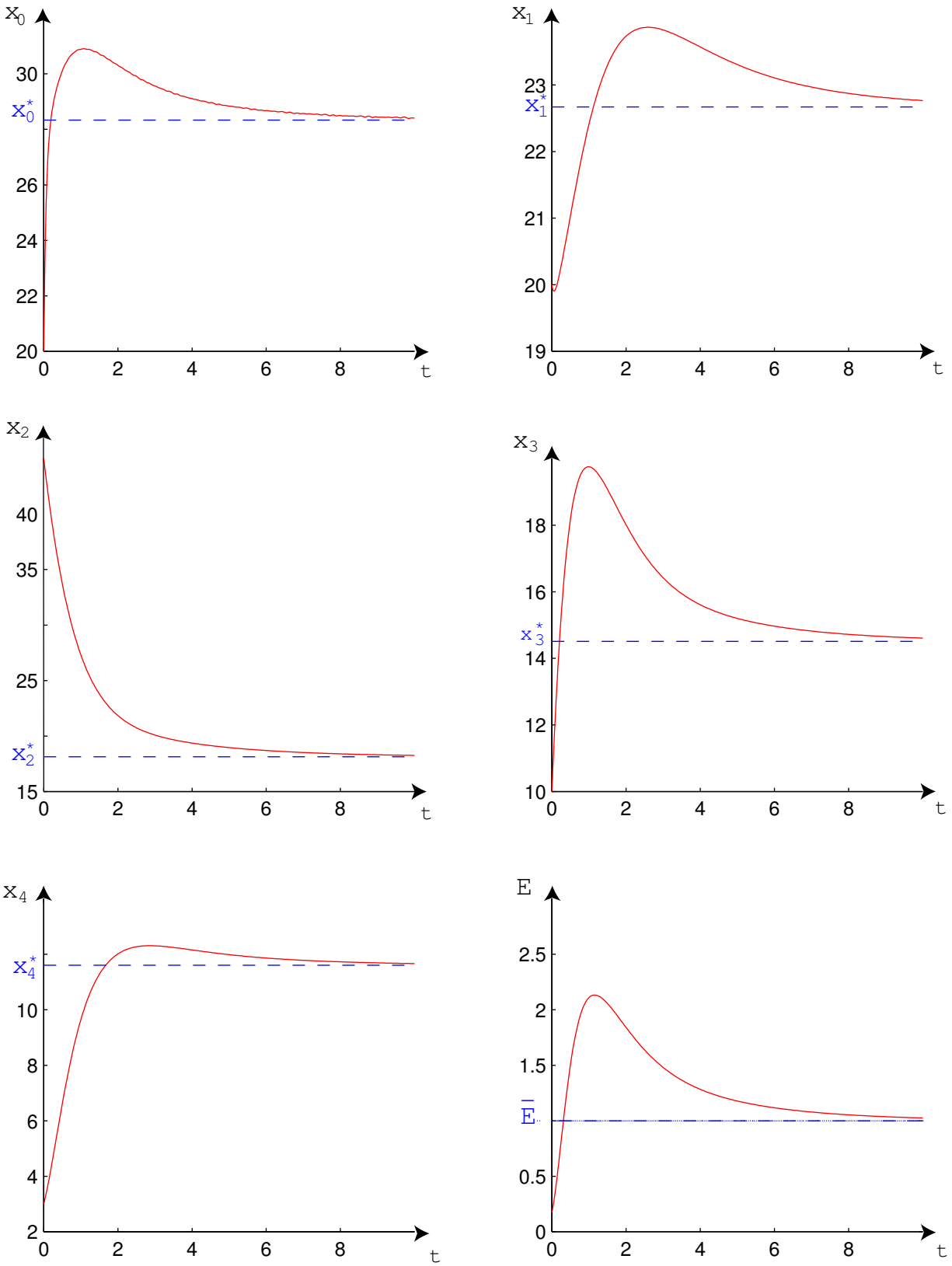

Figure 12: The time evolution of the different stages for the closed-loop system (42-43) with the parameter values given in Table 1 and with initial conditions: $x_0(0) = 20$, $x_1(0) = 20$, $x_2(0) = 45$, $x_3(0) = 10$ and $x_4(0) = 3$.

4 Observability

We consider the standard input-output finite dimensional system:

$$\begin{cases} \dot{x}(t) = X(x(t), u(t)) = X^u(x(t)), \\ y(t) = h(x(t)), \\ x(t) \in M,\ u(t) \in U \subset \mathbb{R}^m,\ y(t) \in \mathcal{Y} \subset \mathbb{R}^q \end{cases} \tag{44}$$

Here M is an open connected subset of $\mathbb{R}^n$. We assume that the vector field X and the map h are C^∞ or analytic (when needed). The point $x(t)$ in M is called the state of the system, $u(t)$ is the input and $y(t)$ is the measurable output of the system. The introduction of y is due to the fact that usually we do not have access to the whole state: we can observe or measure only a part of the actual state of the system.
The class of admissible inputs $\mathcal{U}_{ad} = \{u(.) : t \in \mathbb{R}^+ \mapsto u(t) \in U\}$ is contained in the set of measurable controls with values in U and contains the class of piecewise constant controls with values in U.
We shall use the notation $X_t^{u(.)}(x_0)$ to denote the solution of the differential equation (44) corresponding to the admissible control $u(.) \in \mathcal{U}_{ad}$ and with initial condition x_0: $X_0^{u(.)}(x_0) = x_0$. The corresponding output is $y(x_0, u(.), t) = h\left(X_t^{u(.)}(x_0)\right)$.

Two states $x_1, x_2 \in M$ are said to be *indistinguishable* if $y(x_1, u(.), t) = y(x_2, u(.), t)$ for any admissible control (or input) $u(.) \in \mathcal{U}_{ad}$ and any t for which both sides are defined. 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 of (44) emanating from the state x_1 or by the one emanating from the state x_2.

Example: Consider the single input system with two outputs

$$\begin{cases} \dot{x} &= u \\ y &= h(x) = (\sin x,\ \cos x) \end{cases}$$

where $x \in M = \mathbb{R}$, $u \in \mathbb{R}$, $y = (y_1, y_2) \in \mathbb{R}^2$. We have, for any real x, any integer k and any input u

$$y_1(x + 2k\pi, u(.), t) = \sin\left(x + 2k\pi + \int_0^t u(s)ds\right) = \sin\left(x + \int_0^t u(s)ds\right) = y_1(x, u(.), t).$$

$$y_2(x + 2k\pi, u(.), t) = \cos\left(x + 2k\pi + \int_0^t u(s)ds\right) = \cos\left(x + \int_0^t u(s)ds\right) = y_2(x, u(.), t).$$

Hence the states x and $x + 2k\pi$ are indistinguishable. ■

Two states $x_1,\ x_2 \in M$ $(x_1 \neq x_2)$ are said to be *distinguishable* if there exists an admissible control (or input) $u(.) \in \mathcal{U}_{ad}$ 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.

Consider the single input single output system

$$\begin{cases} \dot{x} &= u \\ y &= h(x) = x^2 \end{cases}$$

where $x \in M = \mathbb{R}$, $u \in \mathbb{R}$, $y \in \mathbb{R}$. Every pair of distinct states x_1 and x_2 are distinguishable: it is sufficient to take $u = c \neq 0$. However, the input $u(t) \equiv 0$ is not an universal input: it does not distinguish the states x and $-x$. ■

The system (44) is *observable* if any pair of distinct initial states $(x_1,\ x_2)$ are distinguishable. The system (44) is *uniformly input observable* (or observable for any input) if for any input $u(.) \in \mathcal{U}_{ad}$ and for any pair of distinct initial states $(x_1,\ x_2)$, there exists a time $t \geq 0$ such that $y(x_1, u(.), t) \neq y(x_2, u(.), t)$.

Example: Bacterial growth in a chemostat.
Consider the following model of the chemostat:

$$\begin{cases} \dot{x} &= \mu(s)\, x - ux \\ \dot{s} &= -k\mu(s)\, x - u(s - s_{in}) \\ y &= x \end{cases} \tag{45}$$

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 a constant. For this system, the input is the flow rate u and the output is usually the concentration $x(t)$. If the map $s \mapsto \mu(s)$ is injective then the chemostat (45) is uniformly input observable. ■

Observability also means that the data of the output $y(t)$ and the input $u(t)$ on any finite time interval $[t_0,\ t_1]$ allow to recover the initial state x_0 and therefore the trajectory starting from this initial state.

4.1 Observability of linear systems

In this section, we study the observability properties of linear systems

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t), \\ x(t) \in \mathbb{R}^n,\ u(t) \in U \subset \mathbb{R}^m,\ y(t) \in \mathbb{R}^q, \\ A,\ B,\ \text{and } C \text{ are respectively } n \times n,\ n \times p \text{ and } m \times n \text{ matrices.} \end{cases} \tag{46}$$

For the linear time-invariant system (46), we have

$$y\,(x_1, u(.), t) = Ce^{tA}x_1 + C\int_0^t e^{(t-s)A}Bu(s)ds\ .$$

Hence,

$$y\,(x_1, u(.), t) - y\,(x_2, u(.), t) = Ce^{tA}(x_1 - x_2).$$

Thus, x_1 and x_2 are indistinguishable if and only if $Ce^{tA}(x_1 - x_2) = 0$ for all $t \geq 0$. By analyticity of $Ce^{tA}x$, this is equivalent to say that all the derivatives of $t \mapsto Ce^{tA}(x_1 - x_2)$ vanish at $t = 0$, i.e,

$$\frac{d^k}{dt^k}Ce^{tA}(x_1 - x_2)\Big|_{t=0} = 0,\ \text{for all } k = 0,\ 1,\ 2, \ldots,$$

which can be written :

$CA^k(x_1 - x_2) = 0,\ \text{for all } k = 0,\ 1,\ 2, \ldots.$

Thanks to Cayley-Hamilton theorem, this is equivalent to :

$C(x_1 - x_2) = CA(x_1 - x_2) = \ldots = CA^{n-1}(x_1 - x_2) = 0.$

Therefore x_1 and x_2 are indistinguishable if and only if

$x_1 - x_2 \in ker\ C\ \cap\ ker\ CA\ \cap\ ker\ CA^2\ \cap \ldots \cap ker\ CA^{n-1}$

We remark that this condition is independent of the input. So the linear system (46) is uniformly input observable if and only if it is observable if and only if the matrix:

$$\mathbf{O}_{(C,A)} = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{pmatrix}$$

is of rank n. In this case we say that the system (46), or the pair $(C,\ A)$, satisfies the Kalman rank condition for observability.

Another interpretation of observability for linear systems: Let us consider now the system (46) with a single output, that is, $y(t) \in \mathbb{R}$ $(q = 1)$. If the system considered is observable (we also say the pair $(C,\ A)$ is observable) then the linear change of coordinates $z = \mathbf{O}_{(C,A)}x$ allows us to write system (46) in the following *observability canonical form* :

$$\begin{cases} \dot{z}(t) = \tilde{A}z(t) + \tilde{B}u(t), \\ y(t) = \tilde{C}z(t), \end{cases} \tag{47}$$

with: $\tilde{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \ldots & -a_{n-1} \end{pmatrix}$ and $\tilde{C} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix}$.

Let us compute the successive time derivatives of the output:

$$\begin{cases} y(t) = \tilde{C}z(t) \\ \dot{y}(t) = \tilde{C}\dot{z}(t) = \tilde{C}\tilde{A}z(t) + \tilde{C}\tilde{B}u(t) \\ \ddot{y}(t) = \tilde{C}\tilde{A}^2z(t) + \tilde{C}\tilde{A}\tilde{B}u(t) + \tilde{C}\tilde{B}\dot{u}(t) \\ \vdots \\ y^{(n-1)}(t) = \tilde{C}\tilde{A}^{n-1}z(t) + \tilde{C}\tilde{A}^{n-2}\tilde{B}u(t) + \tilde{C}\tilde{A}^{n-3}\tilde{B}\dot{u}(t) + \ldots + \tilde{C}\tilde{B}u^{(n-2)}(t) \end{cases}$$

This can be written in a condensed form :

$$\mathbf{O}_{(\tilde{C},\tilde{A})}z(t) = \begin{pmatrix} y(t) \\ \dot{y}(t) - \tilde{C}\tilde{B}u(t) \\ \ddot{y}(t) - \tilde{C}\tilde{A}\tilde{B}u(t) - \tilde{C}\tilde{B}\dot{u}(t) \\ \vdots \\ y^{(n-1)}(t) - \tilde{C}\tilde{A}^{n-2}\tilde{B}u(t) - \tilde{C}\tilde{A}^{n-3}\tilde{B}\dot{u}(t) - \ldots - \tilde{C}\tilde{B}u^{(n-2)}(t) \end{pmatrix}.$$

From this relation we see that, if the functions $t \mapsto y(t)$ and $t \mapsto u(t)$ are known then one can compute the state vector $z(t)$ uniquely since the matrix $\mathbf{O}_{(\tilde{C},\tilde{A})}$ is invertible.

4.2 Observability of nonlinear systems

We go back to the nonlinear system (44). To simplify the notations, we shall consider systems with single output, i.e., $q = 1$ and $y(t) = h(x(t)) \in \mathbb{R}$.

In order to derive an observability condition for (44), we need to recall that a smooth (C^∞) vector field X defined on M operates on $C^\infty(M)$, the set of C^∞ functions $\Phi : M \longrightarrow \mathbb{R}$, by Lie differentiation in the following way:

$$\begin{array}{ccc} C^\infty(M) & \longrightarrow & C^\infty(M) \\ \Phi & \longmapsto & X.\Phi \end{array}$$

with

$$X.\Phi\ (x) = \frac{d}{dt}\Big(\Phi\ \big(X_t(x)\big)\Big)\Big|_{t=0}$$

In the coordinate system $(x_1, \ldots, x_n)$, let $\langle .\ ,\ . \rangle$ denote a scalar product and $\nabla\Phi$ the gradient of Φ in these coordinates. If

$$X(x) = \left(\sum_{i=1}^{n} X_i \partial/\partial x_i\right)(x) = \begin{pmatrix} X_1(x_1, \ldots, x_n) \\ X_2(x_1, \ldots, x_n) \\ \vdots \\ X_n(x_1, \ldots, x_n) \end{pmatrix},$$

then

$$X.\Phi\ (x) = \langle \nabla\Phi(x)\ ,\ X(x) \rangle = \sum_{i=1}^{n} X_i(x) \frac{\partial \Phi}{\partial x_i}(x)$$

The function $X.\Phi$ is called the *Lie derivative* of Φ along the vector field X. It is also denoted $L_X\Phi$. For a given positive integer $k > 0$, the Lie derivative of order k of Φ along X is defined by induction as follows:

$$X^k.\Phi = X.(X^{k-1}.\Phi)\ .$$

Example: $\dot{x} = X(x) = Ax$, $y = h(x) = Cx$. Here, we have $X_t(x_0) = e^{tA}x_0$. Hence $X.h(x_0) = \frac{d}{dt}\Big(Ce^{tA}x_0\Big)\Big|_{t=0} = CAx_0$ and it is easy to see that $X^k.h(x_0) = CA^kx_0$. ■

The **observation space** of (44) $\mathcal{O}$ is the linear space (over $\mathbb{R}$) of functions on M containing the observation function h and which is closed under Lie differentiation by all elements of $\mathcal{F} = \{X^u,\ u \in U\}$. ($\mathcal{F}$ is just the set of vector fields corresponding to constant controls).
It can be proved that $\mathcal{O}$ can be defined as the set of all the linear combination of all repeated Lie derivatives of functions of the form $X^{u_k} \dots X^{u_2}X^{u_1}.h$. i.e.:

$$\mathcal{O} = span_{\mathbb{R}}\{(X^{u_l})^{k_l} \dots (X^{u_2})^{k_2}.(X^{u_1})^{k_1}.h \ :\ l \geq 0,\ \ u_1,\ \dots\ , u_l\ \in U,\ k_i = 0, 1, 2, \dots\}.$$

For analytic systems, the observability is equivalent to the fact that the observability space $\mathcal{O}$ separates the points of M.

Remark: The observation space $\mathcal{O}$ contains the output function and all derivatives of the output function along the system trajectories. In particular, for a system without input

$$\dot{x}(t) = X(x(t)),\ \ y(t) = h(x(t)),$$

$\mathcal{O}$ is constructed by taking $y = h(x)$ together with all repeated time derivatives $\dot{y} = X.h(x)$, $\ddot{y} = X^2.h(x)$, ...

For the linear system (46), the observation space is generated by the functions

$$Cx,\ CAx, \dots,\ CA^{n-1}x.$$

Remark: We have seen that the observability of linear systems does not depend on the input. This is no more true for nonlinear systems as it can be shown by the following example:

$$\begin{cases} \dot{x} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} x + u \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} x, \\ y = x_1. \end{cases} \tag{48}$$

Here $x = (x_1, x_2) \in M = \mathbb{R}^2$, $u \in U = \{0, 1\}$. This is a very simple nonlinear system : it is a bilinear system. This system is observable, because with the input $u(t) \equiv 1$, we obtain an observable linear system $\dot{x} = Ax$, $y = Cx$, with $A = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 0 \end{pmatrix}$. It is easy to see that the Kalman rank condition for observability is satisfied. One also can remark that it is possible to reconstruct the state $(x_1(t), x_2(t))$ from $y(t) = x_1(t)$ and $\dot{y}(t) = x_1(t)+x_2(t)$. Hence, the input $u(t) \equiv 1$ distinguish every pair of distinct initial states x^0 and $\tilde{x}^0$ and this proves that the system (48) is observable. However this system is not uniformly input observable because the input $u(t) \equiv 0$ does not distinguish states x^0 and $\tilde{x}^0$ satisfying: $x_1^0 = \tilde{x}_1^0$ and $x_2^0 \neq \tilde{x}_2^0$.

4.3 Examples from life support systems

Leslie type systems: Models describing the growth of a stage structured population are usually represented by Leslie-type systems :

$$\begin{cases} \dot{x}_1 &= F_1(x_1, x_2, \dots, x_n, u) \\ \dot{x}_2 &= F_2(x_1, x_2, u) \\ \vdots & \\ \dot{x}_{n-1} &= F_{n-1}(x_{n-2}, x_{n-1}, u) \\ \dot{x}_n &= F_n(x_{n-1}, x_n, u) \end{cases} \tag{49}$$

In these models, all the variables are positive, x_1 represents the youngest stage (often eggs). The function F_1 describes the so-called recruitment function, i.e., eggs laying from the other stages. The function $x_{i-1} \mapsto F_i(x_{i-1}, ., .)$ corresponds to the transfer from stage $i-1$ to stage i, and it is often an increasing function. The fishery model (40) is a particular example of Leslie-type systems (49) where the input is the fishing effort.
Suppose we measure only the last stage (the oldest), that is the output of the system (49) is $y = x_n$. This system is observable for any input. Indeed, suppose that for two initial states x and $\bar{x}$ we have the same output, i.e., $x_n(t) = \bar{x}_n(t)$, for all $t \geq 0$ then by differentiation we get:

$$F_n(x_{n-1}(t), x_n(t), u(t)) = F_n(\bar{x}_{n-1}(t), \bar{x}_n(t), u(t)), \ \forall t \geq 0.$$

Since $\partial F_n / \partial x_{n-1}$ is positive, the map $x_{n-1} \mapsto F_n(x_{n-1}, x_n, u)$ is injective. Hence, the above equality (together with $x_n(t) \equiv \bar{x}_n(t)$) implies that $x_{n-1}(t) \equiv \bar{x}_{n-1}(t)$. Once again, by differentiation of this equality we get:

$$F_{n-1}(x_{n-2}(t), x_{n-1}(t), u(t)) = F_{n-1}(\bar{x}_{n-2}(t), \bar{x}_{n-1}(t), u(t)), \ \forall t \geq 0.$$

Thanks to the same argument ($x_{i-1} \mapsto F_i(x_{i-1}, x_i, u)$ is monotone), we deduce that $x_{n-2}(t) \equiv \bar{x}_{n-2}(t)$. And so on, we show that for all $t \geq 0$,

$$\begin{cases} x_n(t) &= \bar{x}_n(t), \\ x_{n-1}(t) &= \bar{x}_{n-1}(t), \\ \vdots & \\ x_1(t) &= \bar{x}_1(t). \end{cases}$$

Hence, $y(x, u(.), t) \equiv y(\bar{x}, u(.), t)$ implies that $x = \bar{x}$ which proves that the Leslie-type system (49) is uniformly input observable.

Trophic chains: Models describing trophic chains represent the dynamics of an ecosystem from nutrient (x_1), phytoplankton (x_2), ... to higher levels such as fish (x_n). They can be written as

$$\begin{cases} \dot{x}_1 = F_1(x_1, x_2, u) \\ \dot{x}_i = F_i(x_{i-1}, x_i, x_{i+1}, u) \quad \text{for } i \in \{2, ..., n-1\} \\ \dot{x}_n = F_n(x_{n-1}, x_n, u). \end{cases} \tag{50}$$

The functions F_i are generally monotone functions of the variable x_{i+1}, i.e., $x_{i+1} \mapsto F_i(.,.,x_{i+1},.)$ is monotone. These systems are uniformly input observable if one can measure on-line the first compartment of the chain or the last one, i.e., $y = x_1$ or $y = x_n$.

5 Observers

We consider an input-output nonlinear system described by

$$\begin{cases} \dot{x} = f(x,u) \\ y = h(x), \end{cases} \tag{51}$$

where $x(t) \in \mathbb{R}^n$ is the state of the system at time t, $u(t) \in U \subset \mathbb{R}^m$ is the input and $y(t) \in \mathbb{R}^q$ is the measurable output of the system.
Here, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ are smooth functions.
We have seen that the observability of the system (51) means that it is theoretically possible to recover the state $x(t)$ from the input $u(t)$ and the output $y(t)$ together with their repeated time derivative $\dot{u},\ \ddot{u}, \ldots, \dot{y},\ \ddot{y}, \ldots, y^{(k)}, \ldots$ along the solution of the system. However, in practice, the use of the derivatives may not give sufficiently accurate performance, especially in the case of noisy measurements as it can be illustrated by the following example:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = u \\ y = x_1 \end{cases} \tag{52}$$

This system is observable for any input. Moreover, the state can be completely recovered because x_1 is measured and one can deduce x_2 and x_3 by $x_2 = \dot{y}$ and $x_2 = \ddot{y}$. However, if the measurement of x_1 is corrupted by a sinusoidal disturbance with a small amplitude then the error on x_2 and x_3 (computed according to the above rules) can become very large. For instance, if $y(t) = x_1(t) + \epsilon \sin(100t)$ then $\dot{y}(t) = x_2(t) + 100\,\epsilon \cos(100t)$ and $\ddot{y}(t) = x_3(t) - 10000\,\epsilon \sin(100t)$.

To avoid the use of the time derivatives of the output and at the same time to counter the induced effects of disturbances, another tool has been developed to reconstruct the state from available outputs. This tool is called *an observer* or *a software sensor*. An observer for the the system (51) is a dynamical system whose inputs are the inputs and outputs of the system (51), which produces an estimate $\hat{x}(t)$ of the state $x(t)$ such that $x(t) - \hat{x}(t) \to 0$ as $t \to +\infty$ and the estimation error must remain small if it starts small.

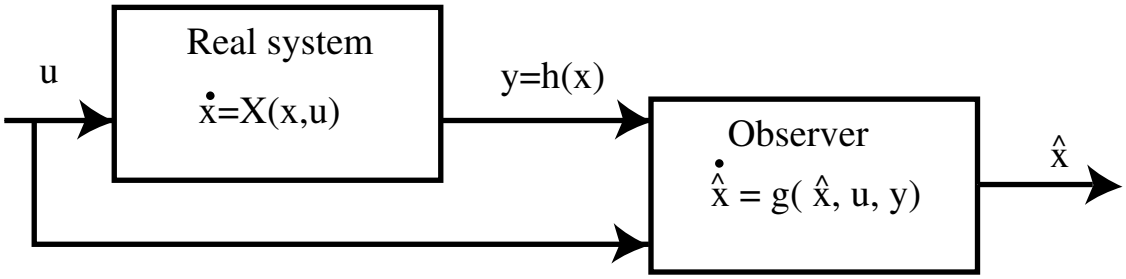

Figure 13: An observer.

A common form for the observers is the following

$$\dot{\hat{x}}(t) = g(\hat{x}(t), y(t), u(t)), \tag{53}$$

where g is smooth. In practice, an on-line estimation of the state is obtained by choosing an initial condition $\hat{x}(0)$ and integrating equation (53) on a computer. We must notice here that it is not possible to do the same thing with the system (51) (which is supposed to model a real system) because we do not know the initial condition $x(0)$.
Let $e(t) = x(t) - \hat{x}(t)$ be the estimation error. The error equation is then given by

$$\dot{e} = f(x, u) - g(\hat{x}, y, u). \tag{54}$$

If the system (53) is an observer for the system (51), then $e = 0$ must be an equilibrium for (54) which implies that $g(x, h(x), u) = f(x, u)$ for all x and all admissible inputs u. Furthermore the null solution $e(t) \equiv 0$ of (54) has to be globally asymptotically stable. If it is only locally asymptotically stable then the dynamical system (53) is a local observer for the system (51). System (53) is said to be an *exponential observer* if the estimation error decreases at an exponential rate, that is, there exist positive constants c and a in such a way that the solutions $x(t)$ and $\hat{x}(t)$ of (51) and (53) satisfy for any initial conditions $x(0)$ and $\hat{x}(0)$

$$\|x(t) - \hat{x}(t)\| \leq c \, \|x(0) - \hat{x}(0)\| \, e^{-a\,t} \ , \ \forall t > 0.$$

To construct an observer for a given system, one has to find the "good" function g that satisfy the above conditions. For linear systems this problem has been solved by the Luenberger observer. For nonlinear systems, there is no "universal" solution but several methods have been developed for some classes of systems. Usually, the simplest and the most natural way to construct an observer for the system (51) is to take a copy of it and to add a corrective term that depends on the difference $h(\hat{x}(t)) - y(t)$, for instance

$$g(\hat{x}, y, u) = f(\hat{x}, u) + K(\hat{x}) \ (h(\hat{x}) - y) \, .$$

5.1 Observers for linear systems

We consider an observable linear system

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t), \\ x(t) \in \mathbb{R}^n, \ u(t) \in U \subset \mathbb{R}^m, \ y(t) \in \mathbb{R}^q, \\ A, \ B, \ \text{and} \ C \ \text{are respectively} \ n \times n, \ n \times m \ \text{and} \ q \times n \ \text{matrices}. \end{cases} \tag{55}$$

An observer (called *Luenberger Observer*) for this system is

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K\Big(y(t) - C\hat{x}(t)\Big) \tag{56}$$

where the $n \times q$ matrix K is to be computed. The observer is an n-dimensional system with state vector $\hat{x}(t)$. The inputs of the observer consist of a copy of the inputs of the original system (55) and the measurements $y(t)$ available from the original system (55). The estimation error $e(t) = x(t) - \hat{x}(t)$ is governed by the differential equation

$$\dot{e}(t) = Ae(t) - K\Big(y(t) - C\hat{x}(t)\Big) = Ae(t) - K\Big(Cx(t) - C\hat{x}(t)\Big) = (A - KC)e(t)\ .$$

System (55) is assumed to be observable, so the pair $(A,\ C)$ is observable and this is equivalent to say that the pair $(A^T,\ C^T)$ (^{T} = transpose) is controllable. For a pair of controllable matrices, we can apply the *Pole-Shifting Theorem* which says that given any $nth-$order real polynomial $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_0$ there is a real matrix F such that the matrix $A^T + C^T F$ has $p(\lambda)$ as its characteristic polynomial. In other words, for any given set $\mathcal{S} = \{\alpha_1, \ldots \alpha_n\}$ of n complex numbers satisfying $z \in \mathcal{S} \Rightarrow \bar{z} \in \mathcal{S}$, it is possible to find a matrix F in such a way that the spectrum of $A^T + C^T F$ is $\sigma(A^T + C^T F) = \mathcal{S}$.
Since the spectrum of a real matrix M and the one of its transpose M^T are equal, we have $\sigma(A^T + C^T F) = \sigma\Big((A^T + C^T F)^T\Big) = \sigma(A + F^T C)$. In particular, there exists a matrix F such that all the eigenvalues of $(A + F^T C)$ are with negative real part. Therefore, if we take $K = -F^T$ then the estimation error satisfies

$$\|e(t)\| \leq c\, e^{-\alpha t},\ \text{where } c > 0,\ \alpha > 0,\ \text{and } \alpha = \max_{\lambda \in \sigma(A-KC)} |Re(\lambda)|.$$

It follows that the Luenberger observer (56) is an exponential observer for the system (55). Moreover the rate of convergence can be arbitrary chosen.

Example: Consider a Leslie type system for the dynamic of a three stage structured population

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -\alpha_1 & 0 & \mu_3 \\ \beta_1 & -\alpha_2 & 0 \\ 0 & \beta_2 & -\alpha_3 \end{pmatrix}}_{A} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}.$$

The entries of the Leslie matrix A are assumed to be independent of the environment and we suppose that we observe only the last stage, that is, the output of the system is $y(t) = x_3(t)$. Here, $C = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$. The transfer from stage $i-1$ to stage i being positive, i.e, $\beta_1 > 0$ and $\beta_2 > 0$, the Kalman observability rank condition is fulfilled. Therefore the following system is an exponential observer (or software sensor) for the above Leslie system:

$$\begin{pmatrix} \dot{\hat{x}}_1(t) \\ \dot{\hat{x}}_2(t) \\ \dot{\hat{x}}_3(t) \end{pmatrix} = \begin{pmatrix} -\alpha_1 & 0 & \mu_3 \\ \beta_1 & -\alpha_2 & 0 \\ 0 & \beta_2 & -\alpha_3 \end{pmatrix} \begin{pmatrix} \hat{x}_1(t) \\ \hat{x}_2(t) \\ \hat{x}_3(t) \end{pmatrix} + (y(t) - \hat{x}_3(t)) \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}.$$

The gain matrix $K = \begin{pmatrix} k_1 & k_2 & k_3 \end{pmatrix}^T$ can be selected in order to force the eigenvalues of $A-KC$ to have desired values. For instance, if one wants to have $\sigma(A-KC) = \{-3,\ -2,\ -1\}$, then the coefficients of the matrix K are:

$$\begin{cases} k_1 = -\dfrac{-6 + 11\,\alpha_1 - 6\,{\alpha_1}^2 + {\alpha_1}^3 - \beta_1\,\beta_2\,\mu_3}{\beta_1\,\beta_2}, \\ k_2 = -\dfrac{-11 + 6\,\alpha_1 - {\alpha_1}^2 + 6\,\alpha_2 - \alpha_1\,\alpha_2 - {\alpha_2}^2}{\beta_2}, \\ k_3 = 6 - \alpha_1 - \alpha_2 - \mu_3. \end{cases}$$

The corresponding estimation error satisfies for all initial conditions $x(0)$ and $\hat{x}(0)$ (recall that $x(0)$ is unknown but $\hat{x}(0)$ can be chosen by the user) and any positive time t

$$\|e(t)\| \leq c\|x(0) - \hat{x}(0)\|\, e^{-3t}.$$

Discrete-time systems: Now we consider an $n-$dimensional discrete-time linear system

$$\begin{cases} x(k+1) = Ax(k) + Bu(k), \\ y(k) = Cx(k). \end{cases}$$

If the pair (C, A) is observable then, by the Pole-Shifting Theorem, it is possible to find a matrix K in such a way that $A - KC$ is a *nilpotent* matrix, i.e, $(A-KC)^n = 0$. The corresponding observer $\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + K\Big(y(k) - C\hat{x}(k)\Big)$ is called a *deadbeat observer*. It has the particularity that the estimate becomes equal to the state after n steps.

5.2 Some nonlinear observers

The construction of nonlinear observers attracted much attention during the last decades. Efficient methods have been designed by many authors. For lack of space we shall only expose the high-gain constructions developed recently by J.P. Gauthier, H. Hammouri, I. Kupka and S. Othman. To avoid complex calculus, we consider only single output nonlinear system.

5.2.1 System with no input

Let us consider an observable single output system

$$\begin{cases} \dot{x} = X(x), \\ y = h(x), \\ x \in \mathbb{R}^n,\ y \in \mathbb{R}. \end{cases} \tag{57}$$

The output function h together with its first $n-1$ derivatives along the vector field X allow to define the following map

$$\begin{cases} \Phi: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & x & \longmapsto & \Phi(x) = (h(x), X.h(x), \ldots, X^{n-1}.h(x))^T \end{cases}$$

We assume that Φ is a global diffeomorphism. This assumption implies that the state can be recovered from the output and its first $n-1$ time-derivatives. In the coordinates defined by $z = \Phi(x)$, the system (57) is governed by the following differential equation:

$$\begin{cases} \dot{z} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & . & . \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}}_{A} z + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \psi(z) \end{pmatrix} = F(z) \\ y = z_1 = \underbrace{(1, 0, \dots, O)}_{C} z. \end{cases} \tag{58}$$

Assume that ψ is globally Lipschitz. Then, for $\theta \geq 1$ large enough, an exponential observer (a *Luenberger type observer*) for the system (58) is given by the following dynamical system:

$$\dot{\hat{z}} = F(\hat{z}) - \Delta_\theta K(C\hat{z} - y), \tag{59}$$

where

$$\Delta_\theta = \begin{pmatrix} \theta & 0 & \dots & 0 \\ 0 & \theta^2 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & \dots & 0 & \theta^n \end{pmatrix},$$

and K is chosen such that the matrix $A - KC$ has all its eigenvalues with negative real part. This is possible since the pair (C, A) is observable.

Remark: In the equation of the observer (59), the term $\Delta_\theta K$ can be replaced by $S_\theta^{-1} C^T$ with S_θ being the solution of

$$\theta S_\theta + A^T S_\theta + S_\theta A = C^T C.$$

The matrix S_θ can be analytically computed by

$$S_\theta(i,j) = \frac{(-1)^{i+j}}{\theta^{i+j-1}} \frac{(i+j-2)!}{(i-1)!(j-1)!}.$$

5.2.2 Affine control systems

Consider a single input single output (SISO) nonlinear system that can be modeled by

$$\begin{cases} \dot{x} = X(x) + uY(x), \\ y = h(x), \\ x \in \mathbb{R}^n, \ y \in \mathbb{R}, \ u \in \mathbb{R}. \end{cases} \tag{60}$$

If the above map Φ defines a global diffeomorphism then the system (60) can be rewritten:

$$\begin{cases} \dot{z} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & . & . \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}}_{A} z + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \psi(z) \end{pmatrix} + u \begin{pmatrix} g_1(z) \\ g_2(z) \\ \vdots \\ g_{n-1}(z) \\ g_n(z) \end{pmatrix} = F(z) + uG(z) \\ y = z_1 = \underbrace{(1, 0, \dots, O)}_{C} z. \end{cases} \tag{61}$$

If moreover the system is observable for any input then the map G satisfies $g_i(z) = g_i(z_1, \dots, z_i)$. This means that

$$\frac{\partial}{\partial z_j}\Big(YX^{i-1}h(\phi^{-1}(z))\Big) = 0, \text{ for } i = 1, \dots, n-1 \text{ and } j = i+1, \dots, n$$

If ψ and g_i are globally Lipschitz then an exponential observer for the system (61) is, for θ large enough, given by:

$$\dot{\hat{z}} = F(\hat{z}) + uG(\hat{z}) - S_\theta^{-1} C^T (C\hat{z} - y). \tag{62}$$

For the original system (60), the observer is:

$$\dot{\hat{x}} = X(\hat{x}) + uY(\hat{x}) - \left[\frac{\partial \Phi}{\partial x}\right]^{-1}_{x=\hat{x}} S_\theta^{-1} C^T (C\hat{x} - y). \tag{63}$$

For the system (60), it is possible to give another type of observer called the *extended Kalman filter*:

$$\begin{cases} \dot{\hat{z}} = F(\hat{z}) + uG(\hat{z}) - \frac{1}{r} S(t)^{-1} C^T (C\hat{z} - y), \\ \dot{S} = -SQ_\theta S - [A^\star(\hat{z}, u)]^T S - SA^\star(\hat{z}, u) + \frac{1}{r} C^T C, \end{cases} \tag{64}$$

where r, θ are positive real numbers, $Q_\theta = \Delta_\theta Q \Delta_\theta$, Q is a given symmetric positive definite $n \times n$ matrix, $A^\star(\hat{z}, u)$ is the Jacobian matrix of $F(z) + uG(z)$ evaluated at $z = \hat{z}$, i.e,

$$A^\star(\hat{z}, u) = \frac{\partial}{\partial z}(F(z) + uG(z))\Big|_{z=\hat{z}}.$$

Once again, for $\theta \geq 1$ and large enough, the system (64) is an exponential observer for the system (61). The difference with the *Luenberger*-like observer (62) is that the gain $S_\theta^{-1} C^T$ used in the correction term is not constant but it is dynamically computed as a solution of a Ricatti matrix differential equation and hence, takes into account the information appearing at the current time t. Therefore, the Kalman observer is more robust with respect to noise that may affect the measurement output.

Example: The chemostat. We consider the following chemostat model

$$\begin{cases} \dot{x} &= \mu(s)\, x - ux, \\ \dot{s} &= -k\mu(s)\, x - u(s - s_{in}), \\ y &= x. \end{cases} \tag{65}$$

Where $\mu(s) = \dfrac{ms}{a+s}$ is the Monod absorption response of the substrate by the micro-organisms. The domain $\Omega = \{\tilde{x} = (x, s) \in \mathbb{R}^2 \ : \ x > 0,\ s > 0,\ s + kx < s_{in}\}$ is positively invariant under the flow of (65). This can be proved by checking that the vector field is always tangent or pointing inside the boundary of Ω. However, the system is not observable on the closure $\bar{\Omega}$ of Ω because all the initial conditions $(0, s)$ produce the same output $y(t) \equiv 0$ and so, for $s_1 \neq s_2$, the states $(0, s_1)$ and $(0, s_2)$ are indistinguishable whatever the input is. Therefore, we shall choose an open subset Ω_1 of Ω such that the system is observable on $\overline{\Omega_1}$. To this end we assume that $0 < u_{min} \leq u \leq u_{max} < m$. Let d be the positive number defined by $d = \dfrac{au_{min}}{m - u_{max}}$ and c be any positive number satisfying $kc + d < s_{in}$. We, then, define Ω_1 by

$$\Omega_1 = \Omega \cap \{(x, s) : x > c,\ s > d\}.$$

It is easy to see that Ω_1 is positively invariant and that the system is observable on $\overline{\Omega_1}$ for all input satisfying the condition $0 < u_{min} \leq u \leq u_{max} < m$.

We perform the change of coordinates given by

$$\Phi : z_1 = y = x,\ z_2 = \mu(s)x = \frac{ms}{a+s}x.$$

Conversely:

$$\Phi^{-1} :\ x = z_1,\ s = \frac{az_2}{mz_1 - z_2}.$$

The domain Ω is transformed by Φ in

$$\Phi(\Omega) = \{(z_1, z_2) \in \mathbb{R}^2 \ : \ z_1 > 0,\ 0 < z_2 < mz_1,\ mkz_1^2 - kz_1z_2 + (a + s_{in})z_2 - ms_{in}z_1 < 0\},$$

and the domain Ω_1 is transformed by Φ in

$$\Phi(\Omega_1) = \{(z_1, z_2) \in \mathbb{R}^2 \ : \ z_1 > c,\ \frac{dm}{a+d}z_1 < z_2 < mz_1,\ mkz_1^2 - kz_1z_2 + (a+s_{in})z_2 - ms_{in}z_1 < 0\}$$

The sets Ω and $\Phi(\Omega)$ (corresponding to the parameters: $a = 10$, $m = 1$, $s_{in} = 30$, $k = 1$) are drawn in Figure 14, and the sets Ω_1 and $\Phi(\Omega_1)$ are drawn in Figure 15.

The map Φ is bijective from Ω_1 to $\Phi(\Omega_1)$. The Jacobian matrices of Φ and Φ^{-1} are

$$\frac{d\Phi}{d\tilde{x}} = \begin{pmatrix} 1 & 0 \\ \dfrac{ms}{a+s} & \dfrac{max}{(a+s)^2} \end{pmatrix}, \quad \frac{d\Phi^{-1}}{dz} = \begin{pmatrix} 1 & 0 \\ \dfrac{-amz_2}{(mz_1 - z_2)^2} & \dfrac{amz_1}{(mz_1 - z_2)^2} \end{pmatrix},$$

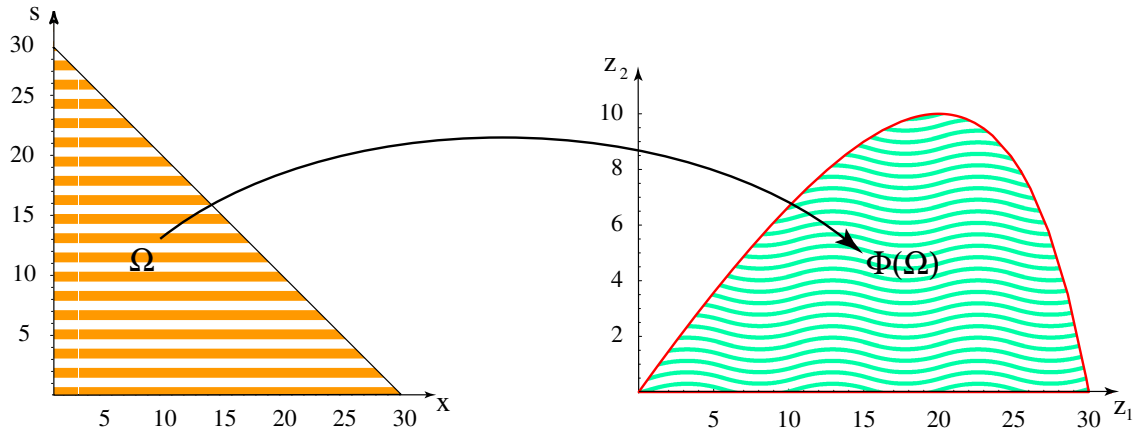

Figure 14: The invariant domain Ω and its image by Φ.

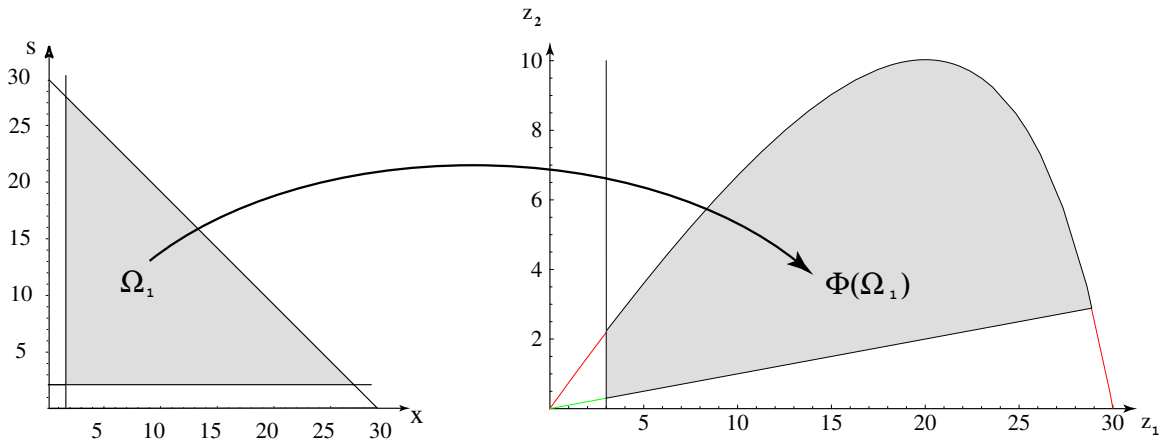

Figure 15: The invariant domain Ω_1 and its image by Φ.

which are invertible for all $(x, s) \in \Omega_1$ and all $(z_1, z_2) \in \Phi(\Omega_1)$. Therefore Φ is a diffeomorphism from Ω_1 to $\Phi(\Omega_1)$ (it is also a diffeomorphism from Ω to $\Phi(\Omega)$). With the new coordinates, the system (65) is defined, for $(z_1, z_2) \in \Phi(\Omega_1)$, by

$$\begin{cases} \dot{z}_1 &= z_2 - uz_1, \\ \dot{z}_2 &= \dfrac{z_2^2}{z_1} - \dfrac{kz_2(mz_1 - z_2)^2}{amz_1} + u\left(-z_2 - \dfrac{(mz_1 - z_2)z_2}{mz_1} + \dfrac{(mz_1 - z_2)^2}{amz_1}s_{in}\right), \\ y &= z_1. \end{cases} \tag{66}$$

We extend this system to the whole $\mathbb{R}^2$ as follows:

$$\begin{cases} \dot{z}_1 &= z_2 - uz_1, \\ \dot{z}_2 &= \psi(z_1, z_2) + ug(z_1, z_2), \\ y &= z_1, \\ (z_1, z_2) \in \mathbb{R}^2, \end{cases} \tag{67}$$

where ψ and g are any C^∞ (or at least continuous) functions that are globally Lipschitz on $\mathbb{R}^2$ and such that their restrictions to $\overline{\Phi(\Omega_1)}$ are:

$$\psi/\overline{\Phi(\Omega_1)} : (z_1, z_2) \mapsto \frac{z_2^2}{z_1} - \frac{kz_2(mz_1 - z_2)^2}{amz_1},$$
$$g/\overline{\Phi(\Omega_1)} : (z_1, z_2) \mapsto -z_2 - \frac{(mz_1 - z_2)z_2}{mz_1} + \frac{(mz_1 - z_2)^2}{amz_1} s_{in}$$

The system (67) has the same form as the system (61) and it satisfies all the required conditions on $\mathbb{R}^2$. Therefore a Luenberger-like observer can be built. To this end, we compute the following matrices

$$S_\theta^{-1} = \begin{pmatrix} 2\theta & \theta^2 \\ \theta^2 & \theta^3 \end{pmatrix}, \quad S_\theta^{-1} C^T = \begin{pmatrix} 2\theta \\ \theta^2 \end{pmatrix}.$$

And then an exponential observer for (67) is

$$\begin{cases} \dot{\hat{z}}_1 &= \hat{z}_2 - u\hat{z}_1 - 2\theta(\hat{z}_1 - y), \\ \dot{\hat{z}}_2 &= \psi(\hat{z}) + ug(\hat{z}) - \theta^2(\hat{z}_1 - y). \end{cases} \tag{68}$$

Therefore, for all $\alpha > 0$, it is possible to find $\theta \geq 1$ such that the solutions of (68-67) satisfy:

$$\exists\, C_1 > 0,\ \forall t \geq 0,\ \forall (\hat{z}(0), z(0)) \in \mathbb{R}^2 \times \mathbb{R}^2,\ \|\hat{z}(t) - z(t)\| \leq C_1 \|\hat{z}(0) - z(0)\| \exp(-\alpha t).$$

Now, let $\tilde{\Phi} : \mathbb{R}^2 \to \mathbb{R}^2$ be a global diffeomorphism such that $\tilde{\Phi}$ and its inverse $\tilde{\Phi}^{-1}$ are globally Lipschitz and such that the restriction of $\tilde{\Phi}$ to the compact set $\overline{\Omega_1}$ is equal to Φ, that is, $\tilde{\Phi}(\tilde{x}) = \Phi(\tilde{x})$ for all $\tilde{x} \in \overline{\Omega_1}$. An estimation of the state vector $\tilde{x}(t) = (x(t), s(t))$ is then given by:

$$\hat{\tilde{x}}(t) = (\hat{x}(t), \hat{s}(t)) = \tilde{\Phi}^{-1}(\hat{z}_1(t), \hat{z}_2(t)) = \tilde{\Phi}^{-1}(\hat{z}(t)), \tag{69}$$

where $\hat{z}(t)$ is given by the differential system (68). This estimation converges exponentially to the real state $\tilde{x}(t)$. Indeed, let K_1 and K_2 be the Lipschitz constants of respectively $\tilde{\Phi}$ and $\tilde{\Phi}^{-1}$, then we have

$$\begin{aligned} \forall t \geq 0,\ \|\hat{\tilde{x}}(t) - \tilde{x}(t)\| &= \|\tilde{\Phi}^{-1}(\hat{z}(t)) - \tilde{\Phi}^{-1} z(t)\| \\ &\leq K_2 \|\hat{z}(t) - z(t)\| \leq K_2 C_1 \|\hat{z}(0) - z(0)\| \exp(-\alpha t) \\ &\leq K_2 K_1 C_1 \|\tilde{x}(0) - \tilde{x}(0)\| \exp(-\alpha t). \end{aligned}$$

Remark: Another way (which is less rigorous than the above one and which is often used for the simulations purpose) to compute the estimations $(\hat{x}(t), \hat{s}(t))$ is just to apply the formula (63) without extending the system to the whole space $\mathbb{R}^2$. To this end we need to compute the following matrices

$$\left[\frac{\partial \Phi}{\partial \tilde{x}}\right]^{-1}_{\tilde{x}=\hat{\tilde{x}}} = \begin{pmatrix} 1 & 0 \\ \dfrac{-\hat{s}(a+\hat{s})}{a\hat{x}} & \dfrac{(a+\hat{s})^2}{am\hat{x}} \end{pmatrix} \quad \left[\frac{\partial \Phi}{\partial \tilde{x}}\right]^{-1}_{\tilde{x}=\hat{\tilde{x}}} S_\theta^{-1} C^T = \begin{pmatrix} 2\theta \\ -2\dfrac{\hat{s}(a+\hat{s})\theta}{a\hat{x}} + \dfrac{(a+\hat{s})^2\theta^2}{am\hat{x}} \end{pmatrix}.$$

The observer for the chemostat is then given by:

$$\begin{cases} \dot{\hat{x}} &= \mu(\hat{s})\,\hat{x} - u\hat{x} - 2\theta(\hat{x}-x), \\ \dot{\hat{s}} &= -k\mu(\hat{s})\,\hat{x} - u(\hat{s}-s_{in}) - \left(-2\dfrac{\hat{s}(a+\hat{s})\theta}{a\hat{x}} + \dfrac{(a+\hat{s})^2\theta^2}{am\hat{x}}\right)(\hat{x}-x). \end{cases} \tag{70}$$

In Figure 16, we present a simulation with the parameter values: $a = 10$, $m = 1$, $s_{in} = 30$, $k = 1$ and with a constant input $u = 0.1$. We have chosen $\theta = 7$. The first figure represents the evolution of, respectively, the biomass x and its estimation $\hat{x}$. The second represents the evolution of, respectively, the substrate s and its estimation $\hat{s}$. The curves show the convergence of the estimations provided by the observer (70) to the states of the system (65) and it can be noted that the convergence is quite fast.

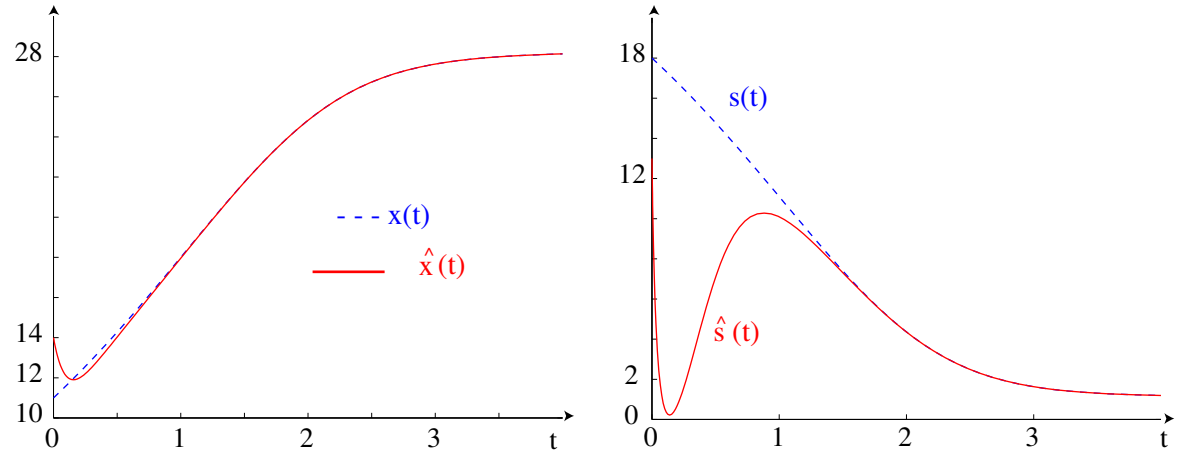

Figure 16: The convergence of the estimations $\hat{x}(t)$ and $\hat{s}(t)$ to the real states $x(t)$ and $s(t)$.

5.2.3 Observers for a class of non-affine control systems

Consider the following analytic system defined on $\mathbb{R}^n$

$$\begin{cases} \dot{x}(t) = \begin{pmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_i(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, u) \\ \vdots \\ f_i(x_1, x_2, \ldots, x_i, x_{i+1}, u) \\ \vdots \\ f_n(x_1, x_2, \ldots, x_n, u) \end{pmatrix} = f(x,u) \\ \\ y = h(x_1, u) \end{cases} \tag{71}$$

where:

- Each of the map f_i is globally Lipschitz with respect to $(x_1, \ldots, x_i)$ uniformly with respect to x_{i+1} and u, i.e, there exists a constant M_i that does not depend neither on x_{i+1} nor on u such that

  $$\|f_i(x_1, \ldots, x_i, x_{i+1}, u) - f_i(z_1, \ldots, z_i, x_{i+1}, u)\| \leq M_i \|(x_1, \ldots, x_i) - (z_1, \ldots, z_i)\|.$$

- There exist two real numbers α, β, $0 < \alpha < \beta$, such that

$$\alpha \leq |\frac{\partial h}{\partial x_1}| \leq \beta, \quad \alpha \leq |\frac{\partial f_i}{\partial x_{i+1}}| \leq \beta, \text{ for } i = 1, \dots n-1.$$

Let $A(t)$ and $C(t)$ be respectively $n \times n$ and $1 \times n$ time-dependent real matrices:

$$A(t) = \begin{pmatrix} 0 & \phi_2(t) & 0 & \dots & 0 \\ 0 & 0 & \phi_3(t) & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \phi_n(t) \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \; C(t) = \begin{pmatrix} \phi_1(t), 0, \dots, 0 \end{pmatrix}.$$

If the functions $\phi_i(t)$ satisfy

$$\alpha \leq \phi_i(t) \leq \beta, \text{ for } i = 1, \dots n-1.$$

Then, there is a real $\lambda > 0$, a vector $K \in \mathbb{R}^n$, and a symmetric positive definite $n \times n$ matrix S, depending only on α and β, such that:

$$\big(A(t) - KC(t)\big)^T S + S\big(A(t) - KC(t)\big) \leq -\lambda Id.$$

The above vector K allows to construct a Luenberger-like observer for the system (71) as follows:

$$\dot{\hat{x}} = f(\hat{x}, u) - \Delta_\theta K(h(\hat{x}_1, u) - y), \tag{72}$$

where $\theta \geq 1$ and $\Delta_\theta = \begin{pmatrix} \theta & 0 & \dots & 0 \\ 0 & \theta^2 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & \dots & 0 & \theta^n \end{pmatrix}$.

This observer is an exponential one. Moreover, it is possible to choose the convergence rate, that is, for any $\alpha > 0$, one can find θ large enough in such a way that for any initial conditions $(x_0, \hat{x}_0)$, the corresponding solutions $x(t)$ of the system (71) and $\hat{x}(t)$ of the observer (72) satisfy

$$\forall t \geq 0, \; \|\hat{x}(t) - x(t)\| \leq P(\alpha) e^{-\alpha t} \|\hat{x}_0 - x_0\|,$$

where P is some polynomial of degree n.

Remark: We have seen that the above observer constructions are made under the assumption that some functions have to be globally Lipschitz. This is a very restrictive condition. However, for real systems, the state space (or at least the set of interest) is often a bounded connected open subset Ω of $\mathbb{R}^n$ and so the global Lipschitz condition is met on $\bar{\Omega}$ the closure of Ω. Therefore the functions considered can be extended to the whole $\mathbb{R}^n$ by smooth and globally Lipschitzian functions on $\mathbb{R}^n$.

5.2.4 Asymptotic observers

For all the above observers, it is possible to assign an arbitrarily (fast) exponential rate of convergence of the error estimation. However, they require the full knowledge of the structure of the model, i.e, the function $X(x,u)$ has to be exactly known and the system has to satisfy some strong observability conditions. When one of the two above conditions is not satisfied, it is still possible in some situations to construct an estimator of the state of the given system but in general, it will not be possible to choose the convergence rate of the estimate to the real state. To illustrate this, we give just two examples. The first one concerns linear systems that are not completely observable, i.e, the rank of the observability matrix is strictly less than the dimension of the state space. The second example concerns biotechnological processes whose dynamical models are partially known, especially the mathematical structure of the biological kinetics is unknown.

Observers for detectable linear systems: Consider again the linear system

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t), \\ x(t) \in \mathbb{R}^n,\ u(t) \in U \subset \mathbb{R}^m,\ y(t) \in \mathbb{R}^q, \\ A,\ B,\ \text{and } C \text{ are respectively } n \times n,\ n \times m \text{ and } q \times n \text{ matrices}, \end{cases} \tag{73}$$

and assume that the pair $(C,\ A)$ is not observable, that is, the rank of the observability matrix $\mathbf{O}_{(C,A)}$ is less than the dimension n of the state space. In this case the Luenberger observer does not work but one can still built an observer for the system (73) if it is *detectable* which means that the unobservable space corresponds to the stable modes of the matrix A. More precisely let $r = \text{rank } \mathbf{O}_{(C,A)}$, there exists a linear change of coordinates $x = Tz$ under which the equations (73) are transformed into

$$\begin{cases} \dot{z}_1 = A_{11}z_1 + B_1u, \\ \dot{z}_2 = A_{12}z_1 + A_{22}z_2 + B_2u, \\ y(t) = C_1z_1, \end{cases} \tag{74}$$

where $z_1 \in \mathbb{R}^r$, $z_2 \in \mathbb{R}^{n-r}$, A_{11} is $r \times r$, A_{12} is $(n-r) \times r$, A_{22} is $(n-r) \times (n-r)$, B_1 is $r \times m$, B_2 is $(n-r) \times m$ and C_1 is $q \times r$. The pair $(C_1,\ A_{11})$ is observable.
The system is detectable if A_{22} is a stable matrix, that is, all its eigenvalues are with negative real part. An observer for (74) is then given by

$$\begin{cases} \dot{\hat{z}}_1 = A_{11}\hat{z}_1 + B_1u + K(y - C_1\hat{z}_1), \\ \dot{\hat{z}}_2 = A_{12}\hat{z}_1 + A_{22}\hat{z}_2 + B_2u. \end{cases} \tag{75}$$

Indeed, the estimation error satisfies the differential equation

$$\begin{cases} \dot{e}_1 = (A_{11} - KC_1)e_1, \\ \dot{e}_2 = A_{12}e_1 + A_{22}e_2. \end{cases}$$

On the one hand, since the pair $(C_1,\ A_{11})$ is observable, it is possible to choose the matrix K in such a way that the matrix $A_{11} - KC_1$ has all its eigenvalues with negative real part. Therefore $\|e_1(t)\|$ converges to zero. On the other hand, the matrix A_{22} is stable so $\|e_1(t)\|$ converges to zero as well. However, we must notice that one can choose the convergence speed of $e_1(t)$ whereas the convergence rate of $e_2(t)$ is completely determined by the eigenvalues of A_{22}.

An observer for a biotechnological process: Let us consider a more general model of the chemostat:

$$\begin{cases} \dot{x} &= \mu(x,s)\ x - u(t)x, \\ \dot{s} &= -k\mu(x,s)\ x - u(t)(s - s_{in}), \\ y &= x. \end{cases} \tag{76}$$

Here we assume that the concentration of micro-organisms $x(t)$ is measured on-line. It is reasonable to assume that the dilution rate (or the flow rate) $u(t)$, the substrate concentration s_{in} in the inflow and the constant k are known. However the function $\mu(x,s)$ is is often a very complex function of the state of the process which depends on the operating conditions. Its analytical expression is still the subject of intensive investigation. For the same process there exist several possible models in the literature, so, which one to choose? For this reason, it is not possible to use one of the above exponential observers. Therefore another type of estimators has been developed for this kind of systems. These estimators are often called *asymptotic observers*. For the chemostat (76) an asymptotic estimator for the unmeasured substrate concentration $s(t)$ can be built as follows. We perform a linear change of coordinates:

$$\begin{cases} z_1 = x, \\ z_2 = kx + s. \end{cases}$$

With these new coordinates, the system (76) becomes:

$$\begin{cases} \dot{z_1} = \mu(z_1, z_2 - kz_1) - u(t)z_1, \\ \dot{z_2} = -u(t)z_2 + u(t)s_{in}. \end{cases}$$

The advantage of the above change of coordinates is that the dynamic of the unmeasured variable z_2 does not depend on the unknown function $\mu(x,s)$. Therefore a candidate estimator for z_2 can be given by

$$\dot{\hat{z}}(t) = -u(t)\hat{z}(t) + u(t)s_{in}.$$

The dynamic of the estimation error $e(t) = \hat{z}_2(t) - z_2(t)$ is

$$\dot{e} = -u(t)e.$$

If the dilution rate $u(t)$ satisfies the condition $0 < u_{min} \leq u(t)$, for all positive time t, then

$$|e(t)| \leq |e(0)|e^{-u_{min}t},\ \forall t \geq 0.$$

Hence an estimate of the unmeasured substrate concentration $s(t)$ is given by $\hat{s}(t)$ which is computed by

$$\begin{cases} \dot{\hat{z}}(t) = -u(t)\hat{z}(t) + u(t)s_{in} \\ \hat{s}(t) = \hat{z}(t) - kx(t). \end{cases}$$

The convergence of the estimator is exponential but its rate of convergence can not be chosen by the user, it is completely determined by the dilution rate $u(t)$.

Acknowledgements
The author is grateful to Prof. C. Lobry for the opportunity to contribute to the EOLSS and to Dr. P. Adda and Prof. G. Sallet for valuable discussions on the subject of the manuscript. A special thank to the anonymous referee for valuable comments and suggestions which helped the author to improve the presentation of this article.

Bibliography

Bacciotti, A. (1992). *Local stabilizability of nonlinear control systems*, 202 p. Series on Advances in Mathematics for Applied Sciences. 8. Singapore, World Scientific. [A well written and easy to read introduction to the stabilization of control systems].

Bastin G. and Dochain D. (1990). *On-line Estimation and Adaptive Control of Bioreactors*, 379 p. Elsevier Publisher. [Bioreactor models have specific structures which are thoroughly considered in this book. In particular, the theory of asymptotic observers is well developed and applied to bioreactors.]

Bernard O., Sallet G. and Sciandra A. (1998). Nonlinear Observer for a Class of Biological Systems: Application to Validation of a Phytoplanktonic Growth Model. *IEEE Trans. Automat. Control*, 43(8), pp 1056–1067. [This paper deals with the observability of some biological systems like Trophic chains and Leslie-type systems. It also presents an application of the observers theory to a particular biological system in order to estimate some state variables that can not be measured].

Bornard G., Celle-Couenne F. and Gilles G. (1993). Observabilité et observateurs. *Systèmes non linéaires, 1. modélisation - estimation* (ed. A.J. Fossard and D. Normand-Cyrot), pp 177–221. Masson. [This paper (in French) is a readable introduction to the observability and observers theory].

Brockett R.W. (1983). Asymptotic stability and feedback stabilization. in *Differential Geometric Control Theory* (Ed. Brockett R.W., Millmann R.S. and Sussmann H.J.), pp. 181–191. Birkhäuser. [This seminal article gives some necessary conditions for the existence of a stabilizing feedback for a nonlinear system].

Gauthier J.-P., Hammouri H. and Othman S. (1992). A simple observer for nonlinear systems, applications to bioreactors. *IEEE Trans. Automat. Control*, 37(6), pp 875–880. [We have used the construction of exponential observers for affine control systems provided by this seminal

article].

Gauthier, J.-P., and Kupka, I. (2001). *Deterministic Observation Theory and Applications*, Cambridge University Press. [This self contained book presents the theoretical foundations of observability. It provides very general results in the construction of exponential observers for nonlinear systems. This monograph requires a good background in differential geometry].

Jurdjevic, V. (1997). *Geometric control theory*, 492 p. Cambridge Studies in Advanced Mathematics. 52. Cambridge University Press. [This book presents the essential aspects of nonlinear control theory including optimal control].

Nijmeijer H. and Van der Schaft A.J. (1990). *Nonlinear Dynamical Control Systems*, 467 pp. New York, Springer-Verlag. [This book presents many techniques for the study of input-output nonlinear systems].

Touzeau S. and Gouzé J.-L. (1998). On the stock-recruitment relationships in fish population models. *Environmental modelling and assessment*, **3**:87–93. [This article gives a stage-structured model of a fish population].