MATH 3530 – MATH 3820
Lecture Notes

Julien Arino & Stéphanie Portet

January 4, 2009

Foreword

These lecture notes are work in progress and therefore, far from complete. They regroup material taught in MATH 3530 (Mathematical Models in Biology) and MATH 3820 (Introduction to Mathematical Modelling). They are derived from the slides shown in class, but some of the computations done in class do not appear here. The content of these lecture notes far exceeds the content of each of these courses.

We have tried to be mathematically rigorous, while not including too much theory. Therefore, a lot of the proofs of results we use are not shown. This does not mean that these proofs are not important, and those of you who intend to do some serious modelling work in the future should definitely try to get a better understanding of these results. A good modeller should know a fair bit of theory in multiple domains (algebra, analysis, graph theory..) and be able to use these results.

We will try to add a section at the end of each chapter to tell you where to get more information on the models, the underlying biological or physical phenomena, as well as leads of where to seek further mathematical explanations. But so far, this remains work to be done.

Contents

Part I

A few introductory considerations

Chapter 1

Introduction to mathematical modelling

Mathematical modelling is an idealization of real-world problems. It is used to help understand mechanisms. Be careful: a model is **almost never** a completely accurate representation of reality.

1.1 Steps of the modelling process

i) identify the most important processes governing the problem (theoretical assumptions)

ii) identify the state variables (quantities studied)

iii) identify the basic principles that govern the state variables (physical laws, interactions)

iv) express mathematically these principles in terms of state variables (choice of formalism)

v) make sure units are consistent

Once a model is obtained

i) identify and evaluate the values of parameters

ii) identify the type of mathematical techniques required for the analysis of the model

iii) conduct numerical simulations of the model

iv) validate the model: it must represent accurately the real process

v) verify the model: it must reproduce know states of the real process

How to represent a problem:

- static vs dynamic
- stochastic vs deterministic
- continous vs discrete
- homogeneous vs detailed

Formalism: ODE, PDE, DDE, SDE, integral equations, integro-differential equations, Markov Chains, game theory, graph theory, cellular automata, L-systems ....

1.2 Example: biological problems

- ecology (predator-prey system, populations in competition . . .)
- etology
- epidemiology (propagation of infectious diseases)
- physiology (neuron, cardiac cells, muscular cells)
- immunology
- cell biology
- structural biology
- molecular biology
- genetics (spread of genes in a population)

Chapter 2

A gentle introduction to Maple and Matlab

2.1 Maple

2.2 Matlab

The "Mat" in Matlab does not stand for "mathematics", but for "matrix".. all objects in matlab are matrices of some sort! Keep this in mind when using this program. Matlab is a high level *interpreted* programming language:

- a matlab program is typically a set of instructions that are evaluated iteratively;
- most of the work can be done directly from the command line. For convenience, however, it is possible to store these instructions in files, if they are going to be used repeatedly.

2.2.1 Computing iterates

Defining a function from the command line

We want to plot the iterates of some function f. First, we define the function.

```
>> f=inline('r.*x.*(1-x)','x','r')
f =
     Inline function:
     f(x,r) = r.*x.*(1-x)
```

This defines a function (here, with two arguments, x and r), that can then be used:

```
>> f(0.2,3.2)
ans =
    0.5120
```

Defining a function in a .m file

The other way to define a function, which is more convenient when defining elaborate functions, is to store the function in a `.m` file.

";" hides the result on the command line Remark that

```
>> f(0.2,3.2)
ans =
    0.5120
```

but

```
>> f(0.2,3.2);
```

produces no output.

Creating a vector To create a vector, use the command

$$x = \text{first entry} : \text{step} : \text{last entry},$$

or, if entries are a subset of the integers,

$$x = \text{first entry} : \text{last entry}.$$

For example, we want to plot the iterates of the logistic map (see details in Section 3.5), so

```
x=0:0.01:1;
```

Note the ";": otherwise, we get the full 101 elements vector displayed.

What is the size of .. ? As mentioned, in matlab everything is a matrix. For matrix operations, size is important, and it is frequent to make mistakes. To check, `whos` and `size`. `whos` gives a lot of information.

```
>> whos x
  Name      Size                Bytes  Class

  x         1x101                 808  double array
Grand total is 101 elements using 808 bytes
```

Various variables can be listed on the line after `whos`:

```
>> whos x k
  Name      Size                  Bytes  Class

  k         1x1                       8  double array
  x         1x101                   808  double array
Grand total is 102 elements using 816 bytes
```

If no variable name is provided, `whos` returns the whole workspace, i.e., all variables and their size in memory.

`size`, on the other hand, is "attributable". It can be used like this

```
>> size(x)
ans =
     1   101
```

but, since the result is a vector

```
>> [r,c]=size(x)
r =
     1
c =
   101
```

in which case, r and c take the values of the numbers of rows and columns, respectively.

Vectorized functions versus nonvectorized functions Recall that we wrote

```
>> f=inline('r.*x.*(1-x)','x','r')
```

that is, every multiplication sign took the form `.*` instead of `*`. Here, this is needed: we want to use the *vectorized* form of the function, and be able to pass to f a vector instead of a single value. The `.*` form means that the operation is applied to every entry in the vector/matrix. Same exists for `/` and `^`. It is also possible to use the function `vectorize`, which transforms the function into its vectorized equivalent.

The result of using this vectorized form is that f will be applied to every entry of x, and will produce a vector.

Vectorized operations have been optimized in matlab, and are extremely fast. When possible, they should be used instead of loops.

Vectorized vs nonvectorized Define

```
>> f=inline('r.*x.*(1-x)','x','r')
>> g=inline('r*x*(1-x)','x','r')
```

and for simplicity, consider the vector

```
>> x=[1,2];
```

Then

```
>> f(x,3.5)
   g(x,3.5)
ans =
     0    -7

??? Error using ==> inlineeval
Error in inline expression ==> r*x*(1-x)
??? Error using ==> mtimes
Inner matrix dimensions must agree.
```

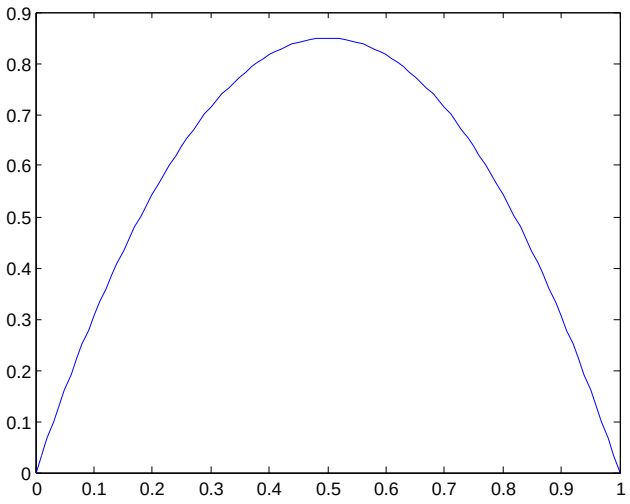
Plotting Basic plotting is very easy. The format is

```
plot(x_axis,y_value)
```

so, for example (with f as defined above),

```
plot(x,f(x,3.4))
```

(here, ";" or not does not matter, as the figure appears in a new window and all that ";" changes is the output in the command window).



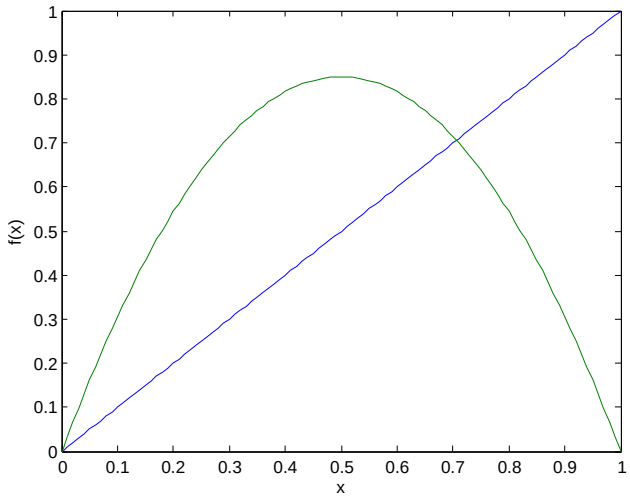
This is a very basic plot. We could want to plot more than one object (for example, the line $y = x$ would be nice).

```
plot(x,x,x,f(x,3.4));
```

Ordering is by pairs: $x_1, f_1(x_1), x_2, f_2(x_2)$. Two elements in a pair **must have** the same number of columns. Different pairs **can have** different numbers of columns. Each element in a given pair can be a point, a vector, a matrix.

We could also want to label the axes.

```
xlabel('x');
ylabel('f(x)');
```



Computing several iterates For the moment, we only have $f(x)$. We want $f^n(x)$, for a given n. Several ways.

- Taking for example $r = 3.5$, use

  ```
  f(f(x,3.5),3.5)
  ```

- The downside to this method is that matlab does not allow to formally define f^n, so tricks have to be used for larger values of n, for example, produce a string containing the command

  ```
  f(f(f(f(f(x,3.5),3.5),3.5),3.5),3.5)
  ```

  and evaluate it. This is feasible but complicated.

- Another method consists in using the result found at the previous step to evaluate the next. We do this now.

Automatic resizing of vectors and matrices We are going to use a very nice feature of matlab: adding elements to a vector, or rows/columns to a matrix, is automatic. Suppose for example that we had defined x as

```
x=0:0.01:0.5;
```

Then

```
x=[x,0.51:0.01:1];
```

would produce the vector x as we had earlier.
Be careful! Note that the command was

```
x=[x,0.51:0.01:1];
```

that is, the old and new entries were separated by a ",". This is *horizontal concatenation*. The command with a ";" tries to add a new row. In our case, we get

```
>> z=[z;0.51:0.01:1]
??? Error using ==> vertcat
All rows in the bracketed expression must have the same
number of columns.
```

because we are trying to add a row of 50 elements to a row of 51 elements. But

```
>> z=[z;0.51:0.01:1.01]
```

works, and gives a 2×51 matrix.

Here, we are going to use the latter form of the command, and add each successive iterate to a solution matrix M. First, define an empty matrix,

```
M=[];
```

Then we need to loop from 1 to n, where n is the iterate that we want.

Loops The command uses the same type of syntax as the creation of a vector: to loop from 4 to 12 by steps of 1,

```
for i=4:12,
   command(s) to be repeated, maybe using the value i
end;
```

whereas to loop by non-unit or non-integer steps, say from 4 to 12 by steps of 1.35,

```
for i=4:1.35:12,
   command(s) to be repeated, maybe using the value i
end;
```

Note that in that case, the last i is equal to 10.75, not 12, since $10.75 + 1.35 = 12.1 > 12$. The same is true when using non-unit steps to create vectors.

Accessing matrix elements Suppose that M is an $m \times n$-matrix. Then

- `M(i,j)` is the element on the ith row and jth column.
- `M(i,:)` is the ith row.
- `M(:,j)` is the jth column.
- `M(end,:)` is the last row of M (`end` is a reserved word which always points to the last valid index in a given matrix dimension).
- `M(:,end)` is the last column of M.
- `M(end,1:10)` are the first 10 entries in the last row of M.
- `M(1:2,3:5)` is the submatrix of M consisting of rows 1 and 2 and columns 3 to 5 of M.

Back to the iterates After some thought, we realize that we will need to go back one iterate. So instead of starting with empty matrix M, fill the first row of M with first iterate, and start at iterate 2.

```
n=10;
r=3.5;
M=f(x,r);
for i=2:n,
    M=[M;f(M(end,:),r)];
end;
plot(x,M);
```

This plots all the iterates to n. The result is a bit crowded, as can be seen in Figure 2.1.

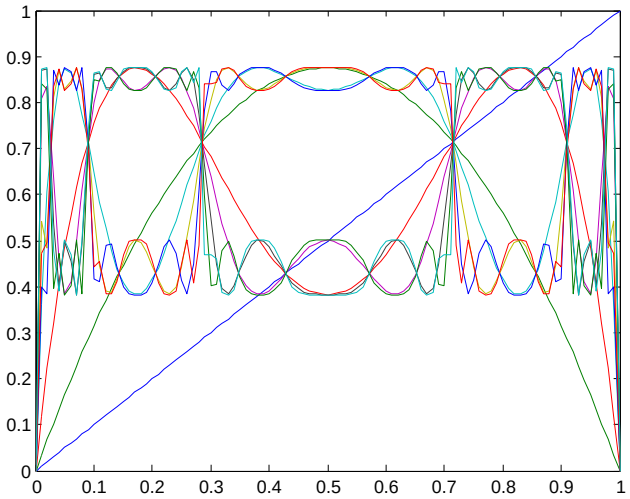


Figure 2.1: Plot of the first 10 iterates of the logistic map.

2.2.2 Numerical simulation of differential equations

We want to compute the numerical solution to the logistic equation, which is given by

$$N' = rN\left(1 - \frac{N}{K}\right), \tag{2.1}$$

with r the intrinsic growth rate of the population and K the carrying capacity (see Chapter 3 for details). There are many ways to do this. We can use, for example,

- matlab
- octave
- scilab
- maple
- mathematica
- many others..

We recommend using matlab, octave or scilab are recommended because of their "philosophy", which is very close to the "natural" way one proceeds with an ode. Note also that these programs are geared toward numerical simulations, and are therefore very efficient in that context.

A brief reminder about Euler's method helps to understand the above remark about the "philosophy" of the programs. The solution to the initial value problem

$$\begin{aligned} x' &= f(t, x) \\ x(t_0) &= x_0 \end{aligned}$$

can be approximated numerically by the following sequence:

$$t_{k+1} = t_k + h$$
$$x_{k+1} = x_k + hf(t_k, x_k)$$

for a time step $h > 0$ and with first term (t_0, x_0). The techniques (a.k.a. "numerical solvers") in matlab are much more advanced, but the idea is the same: approximate the solution to an ODE by using a numerical algorithm that uses information on the "shape" of the vector field. We need two files:

i) a RHS function defining $f(t, x)$

ii) a function or command line statement that "calls" the RHS function with a numerical solver

For the logistic equation (2.1), we could define the following function:

```
function dN=rhs_logistic(t,N,p)
% This function returns the value of dN/dt
% at the point (t,N), using parameters in the
% structure p

dN=p.r*N*(1-N/p.K);
```

which we save in a file called, say, `rhs_logistic.m`. Note that `t` is required in the function arguments even if not used in the RHS function, i.e., even if f is autonomous.

In our code above, the variable `p` is defined as a *structure*. This is a very useful construct in many programming languages. Think of it as a *container*:

```
>> p.K=100;
>> p.r=2;
>> p
p =
    K: 100
    r: 2
```

Pros: `p` is passed to the function as one parameter, instead of a list of parameters. Cons: do not forget `p.` in front of the parameter. We will see later why structures are useful

Once the right hand side function is set up, we need to invoke the numerical solver. The call is of the form (from the help):

```
ode23, ode45, ode113, ode15s, ode23s, ode23t, ode23tb

Solve initial value problems for ordinary differential
equations

Syntax
[T,Y] = solver(odefun,tspan,y0)
[T,Y] = solver(odefun,tspan,y0,options)
[T,Y,TE,YE,IE] = solver(odefun,tspan,y0,options)
```

```
sol = solver(odefun,[t0 tf],y0...)

where solver is one of ode45, ode23, ode113, ode15s,
ode23s, ode23t, or ode23tb
```

Typically, you can use `ode45`

Computing the numerical solution to the logistic We call our solver as follows:

```
tspan=[1790 2000]; %The time span of the solution
IC=3.929;          %The initial condition (in 1790)
p.K=300;           %Set the parameters
p.r=0.5;
[t,N]=ode45(@rhs_logistic,tspan,IC,[],p);
```

(The one before last argument, `[]`, represents the options structure. Here we are not modifying any option, and so pass an empty vector)

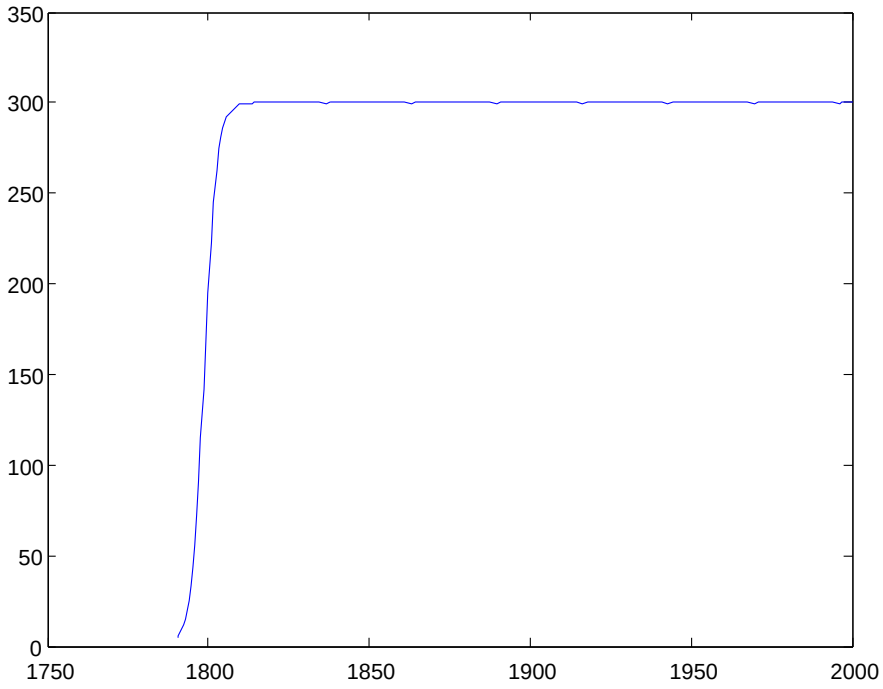
Save this file as, say, `call_solver.m`

After running it, we have a vector `t` of times (covering `tspan`) and a vector `N` of solution [containsverbatim]

Plotting the solution

```
plot(t,N)
```

gives

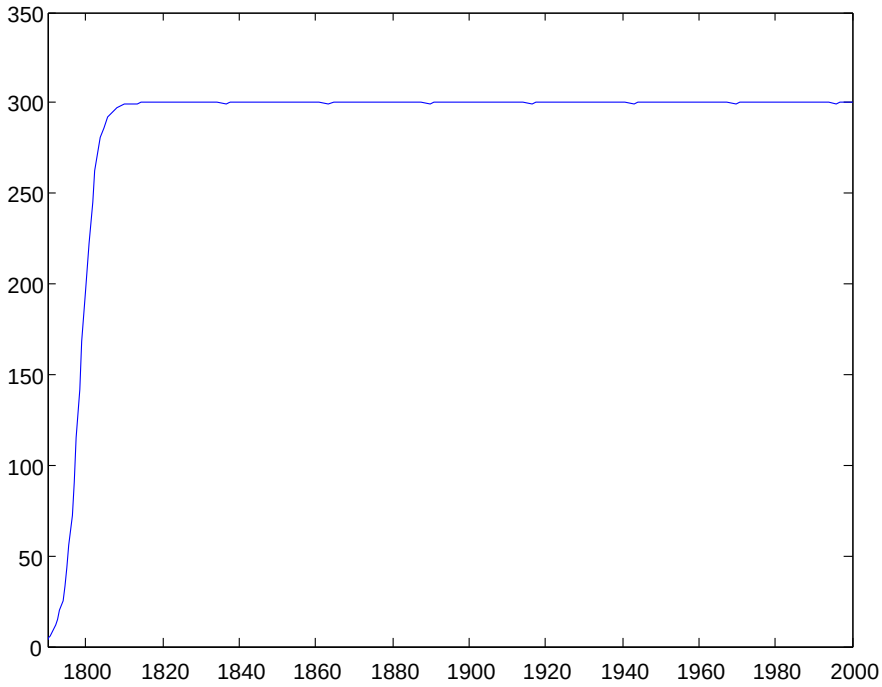


[containsverbatim]

Tightening the x-axis

```
plot(t,N)
xlim([t(1) t(end)])
```

gives



[containsverbatim]

Using Octave The syntax in Octave is almost identical to the matlab syntax. In fact, if you use the additional programs in the `forge` repository, a function `ode45` is defined

However, the functions (in octave) do not implement the use of a parameter by default, so a work-around must be used

Update: as of V3.0 and using ode45, parameters can be passed and the matlab code given before works, with the following little modification:

```
opt=odeset('InitialStep',0.05,'MaxStep',1);
[t,N]=ode45(@rhs_logistic,tspan,IC,opt,p);
```

which makes sure that the time step does not become too large)
[containsverbatim]

Using scilab The syntax in scilab differs a little from matlab, so beware.

```
function ydot=f(t,y);
ydot=y^2-y*sin(t)+cos(t);
endfunction
```

Chapter 3

A single population growth model: the logistic curve/equation/map

The order in which the models are presented in this chapter is different from the rest of the manuscript.

Add PDE logistic

3.1 Objectives

We are given a table with the population census at different time intervals between a date a and a date b, and want to get an expression for the population. This allows us to:

- compute a value for the population at any time between the date a and the date b (**interpolation**),
- predict a value for the population at a date before a or after b (**extrapolation**).

This was studied in a series of papers in the 1920-40's, mainly under the influence of Pearl and Reed [11, 12, 13].

3.2 The data: US census

Year	Population (millions)	Year	Population (millions)
1790	3.929	1850	23.192
1800	5.308	1860	31.443
1810	7.240	1870	38.558
1820	9.638	1880	50.156
1830	12.866	1890	62.948
1840	17.069	1900	75.995
		1910	91.972

Table 3.1: The US population from 1790 to 1910. From Pearl and Reed [11].

Using MatLab (or Octave), create two vectors using commands such as

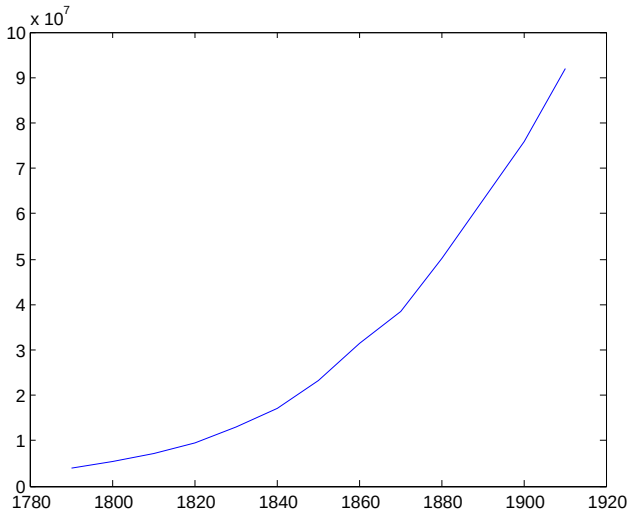
```
t=1790:10:1910;
```

which creates the vector of time points, and

```
P=[3929214,5308483,7239881,9638453,12866020, ...
    17069453,23191876,31443321,38558371,50155783, ...
    62947714,75994575,91972266];
```

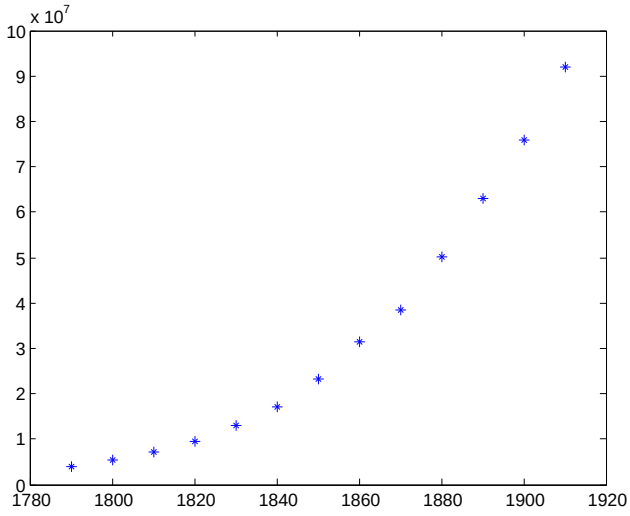
for corresponding population values (`...` indicates that the line continues below). Then plot using

```
plot(t,P);
```



To get points instead of a line, use the command

```
plot(t,P,'*');
```



3.2.1 A quadratic curve?

The curve looks like part of a parabola. So let us use nonlinear regression to fit a curve of the form

$$P(t) = a + bt + ct^2$$

to the data. We proceed as follows. There are 13 data points which we denote (t_k, P_k) for $k = 1, \ldots, 13$. The objective of nonlinear regression is to find the values of a, b, c such that

$$S = \sum_{k=1}^{13} \left(P(t_k) - P_k\right)^2$$

be minimal. This means that the values a, b, c are such that the square of the distance between the the known points (t_k, P_k) and those for corresponding times, $(t_k, P(t_k)) = (t_k, a + bt_k + ct_k^2)$, is minimal. To emphasize the dependence of S on the values of a, b, c, we denote it

$$S(a, b, c) = \sum_{k=1}^{13} \left(a + bt_k + ct_k^2 - P_k\right)^2.$$

We have that $S(a, b, c)$ is minimal if (necessary condition) $\partial S/\partial a = \partial S/\partial b = \partial S/\partial c = 0$, with

$$\begin{aligned}
\frac{\partial S}{\partial a} &= 2\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k) \\
\frac{\partial S}{\partial b} &= 2\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k)t_k \\
\frac{\partial S}{\partial c} &= 2\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k)t_k^2.
\end{aligned}$$

So we want

$$\begin{aligned}
2\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k) &= 0 \\
2\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k)t_k &= 0 \\
2\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k)t_k^2 &= 0,
\end{aligned}$$

which we can simplify,

$$\begin{aligned}
\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k) &= 0 \\
\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k)t_k &= 0 \\
\sum_{k=1}^{13}(a + bt_k + ct_k^2 - P_k)t_k^2 &= 0.
\end{aligned}$$

Rearranging the system, we get

$$\begin{aligned}\sum_{k=1}^{13}(a+bt_k+ct_k^2) &= \sum_{k=1}^{13} P_k\\ \sum_{k=1}^{13}(at_k+bt_k^2+ct_k^3) &= \sum_{k=1}^{13} P_kt_k\\ \sum_{k=1}^{13}(at_k^2+bt_k^3+ct_k^4) &= \sum_{k=1}^{13} P_kt_k^2.\end{aligned}$$

After a bit of tidying up, and emphasizing the fact that the unknowns are a, b, c, we get

$$\begin{aligned}\left(\sum_{k=1}^{13}1\right)a+\left(\sum_{k=1}^{13}t_k\right)b+\left(\sum_{k=1}^{13}t_k^2\right)c &= \sum_{k=1}^{13}P_k\\ \left(\sum_{k=1}^{13}t_k\right)a+\left(\sum_{k=1}^{13}t_k^2\right)b+\left(\sum_{k=1}^{13}t_k^3\right)c &= \sum_{k=1}^{13}P_kt_k\\ \left(\sum_{k=1}^{13}t_k^2\right)a+\left(\sum_{k=1}^{13}t_k^3\right)b+\left(\sum_{k=1}^{13}t_k^4\right)c &= \sum_{k=1}^{13}P_kt_k^2.\end{aligned}$$

So we need to solve the linear system

$$\begin{pmatrix} 13 & \sum_{k=1}^{13}t_k & \sum_{k=1}^{13}t_k^2\\ \sum_{k=1}^{13}t_k & \sum_{k=1}^{13}t_k^2 & \sum_{k=1}^{13}t_k^3\\ \sum_{k=1}^{13}t_k^2 & \sum_{k=1}^{13}t_k^3 & \sum_{k=1}^{13}t_k^4\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix} = \begin{pmatrix}\sum_{k=1}^{13}P_k\\ \sum_{k=1}^{13}P_kt_k\\ \sum_{k=1}^{13}P_kt_k^2\end{pmatrix}.$$

With MatLab (or Octave), getting the values is easy.

- To apply an operation to every element in a vector or matrix, prefix the operation with a dot, hence

  ```
  t.^2;
  ```

  gives, for example, the vector with every element t_k squared.

- Also, the function `sum` gives the sum of the entries of a vector or matrix.

- When entering a matrix or vector, separate entries on the same row by `,` and create a new row by using `;`.

Thus, to set up the problem in the form of solving $Ax = b$, we need to do the following:

```
format long g;
A=[13,sum(t),sum(t.^2);sum(t),sum(t.^2),sum(t.^3);...
sum(t.^2),sum(t.^3),sum(t.^4)];
b=[sum(P);sum(P.*t);sum(P.*(t.^2))];
```

The `format long g` command is used to force the display of digits (normally, what is shown is in "scientific" notation, not very informative here).

Then, solve the system using

```
A\b
```

We get the following output:

```
>> A\b
Warning: Matrix is close to singular or badly scaled.
         Results may be inaccurate. RCOND = 1.118391e-020.

ans =
        22233186177.8195
        -24720291.325476
        6872.99686313725
```

(note that here, Octave gives a solution that is not as good as this one, provided by MatLab).

3.2.2 Checking our results for the quadratic

Thus

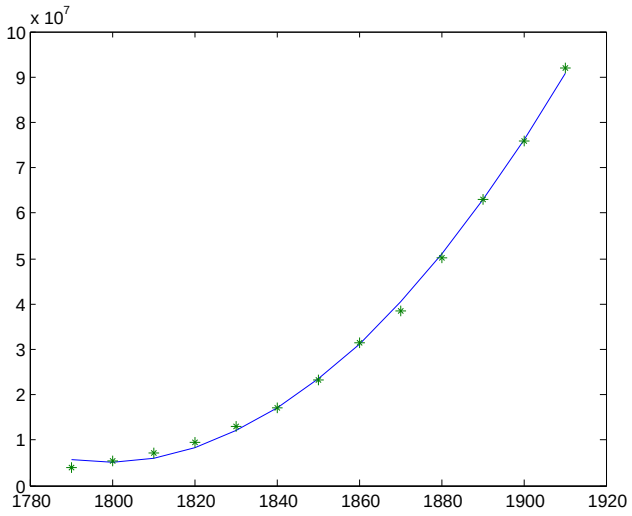
$$P(t) = 22233186177.81 - 24720291.32t + 6872.99t^2$$

To see what this looks like,

```
plot(t,22233186177.81-24720291.32.*t+6872.99.*t.^2);
```

(note the dots before multiplication and power, since we apply this function to every entry of t). In fact, to compare with original data:

```
plot(t,22233186177.81-24720291.32.*t+6872.99.*t.^2,t,P,'*');
```



Now we want to generate the table of values, to compare with the true values and thus compute the error. To do this, we can proceed directly:

```
computedP=22233186177.81-24720291.32.*t+6872.99.*t.^2;
```

We get

```
computedP =
 Columns 1 through 4:
      5633954.39      5171628.52      6083902.03    8370774.90
 Columns 5 through 8:
      12032247.15      17068318.78      23478989.77    31264260.14
 Columns 9 through 12:
       40424129.88      50958598.99      62867667.48    76151335.34
 Column 13:
      90809602.57
```

We can also create an **inline** function

```
f=inline('22233186177.81-24720291.32.*t+6872.99.*t.^2')
f =
     Inline function:
     f(t) = 22233186177.81-24720291.32.*t+6872.99.*t.^2
```

This function can then easily be used for a single value

```
octave:24> f(1880)
ans =     50958598.9969215
```

as well as for vectors.

(Recall that t has the dates; t in the definition of the function is a dummy variable, we could have used another letter-.)

```
octave:25> f(t)
ans =
 Columns 1 through 4:
      5633954.39      5171628.52      6083902.03    8370774.90
 Columns 5 through 8:
      12032247.15      17068318.78      23478989.77    31264260.14
 Columns 9 through 12:
       40424129.88      50958598.99      62867667.48    76151335.34
 Column 13:
      90809602.57
```

Form the vector of errors, and compute sum of errors squared:

```
octave:26> E=f(t)-P;
octave:27> sum(E.^2)
ans =     12186176863781.4
```

Quite a large error (12,186,176,863,781.4), which is normal since we have used actual numbers, not thousands or millions of individuals, and we are taking the square of the error.

To present things legibly, one way is to put everything in a matrix..

```
M=[P;f(t);E;E./P];
```

This matrix will have each type of information as a row, so to display it in the form of a table, show its transpose, which is achieved using the function `transpose` or the operator ′.

```
M'
ans =
 3929214      5633954.39      1704740.39     0.43
 5308483      5171628.52     -136854.47     -0.02
 7239881      6083902.03     -1155978.96    -0.15
 9638453      8370774.90     -1267678.09    -0.13
12866020      12032247.15     -833772.84    -0.06
17069453      17068318.78     -1134.21   -6.644728828e-05
23191876      23478989.77      287113.77     0.01
31443321      31264260.14     -179060.85   -0.00569471832254123
38558371      40424129.88      1865758.88     0.04
50155783      50958598.99      802815.99     0.01
62947714      62867667.48     -80046.51   -0.00127163502018304
75994575      76151335.34      156760.34    0.00206278330494212
91972266      90809602.57     -1162663.42     -0.01
```

Now for the big question... How does our formula do for present times?

```
f(2006)
ans =     301468584.066013
```

Actually, quite well: 301,468,584, compared to the 298,444,215 July 2006 estimate, overestimates the population by 3,024,369, a relative error of approximately 1%.

The US population from 1790 to 2000 (revised numbers)

Year	Population (millions)	Year	Population (millions)
1790	3.929	1900	76.212
1800	5.308	1910	92.228
1810	7.240	1920	106.021
1820	9.638	1930	123.202
1830	12.866	1940	132.164
1840	17.069	1950	151.325
1850	23.192	1960	179.323
1860	31.443	1970	203.302
1870	38.558	1980	226.542
1880	50.156	1990	248.709
1890	62.948	2000	281.421

3.2.3 Other similar approaches – The logistic curve

Pritchett [15] tried

$$P = a + bt + ct^2 + dt^3,$$

using data from 1790 to 1880 (inclusive). Pearl and Reed, in [11], start by comparing the results of [15] with those found by using

$$P(t) = a + bt + ct^2 + d\ln t.$$

They find

$$P(t) = 9,064,900 - 6,281,430t + 842,377t^2 + 19,829,500\ln t.$$

and a cumulative error (S) half of that of Pritchett. They then try

$$P(t) = \frac{be^{at}}{1 + ce^{at}}$$

or

$$P(t) = \frac{b}{e^{-at} + c}. \tag{3.1}$$

They find

$$P(t) = \frac{2,930.3009}{e^{-0.0313395t} + 0.014854}.$$

3.3 The ODE logistic equation

The *logistic curve* (3.1) is the solution to an *ordinary differential equation* called the **logistic equation**. This equation was introduced by Pierre-François Verhulst (1804-1849) [16, 17]. The idea is to represent a population evolving subject to the following effects:

- birth, at the **per capita** rate b,
- death, at the **per capita** rate d,
- competition of individuals with other individuals reduces their ability to survive, resulting in death.

This gives

$$N' = bN - dN - \text{competition}.$$

Competition describes the mortality that occurs when two individuals meet:

- In chemistry, if there is a concentration X of one product and Y of another product, then XY, called **mass action**, describes the number of interactions of molecules of the two products.
- Here, we assume that X and Y are of the same type (individuals). So there are N^2 contacts.
- These N^2 contacts lead to the death of individuals at the rate c.

Therefore, the **logistic** equation is

$$N' = bN - dN - cN^2. \tag{3.2}$$

Rewriting this equation as

$$N' = (b - d)N - cN^2,$$

we see that

- $b - d$ is the rate at which the population increases (or decreases) in the absence of competition. It is called the **intrinsic growth rate** of the population.
- c is the rate of **intraspecific** competition. The prefix **intra** refers to the fact that the competition is occurring between members of the same species, that is, within the species. [We will see later examples of **interspecific** competition, that is, between different species.]

Factor out an N in (3.2), giving

$$N' = \big((b-d) - cN\big)N.$$

This gives us the original interpretation of the logistic equation, since, writing

$$\frac{N'}{N} = (b-d) - cN,$$

we have N'/N, the **per capita growth rate** of N, given by a constant, $b-d$, minus a **density dependent inhibition** factor, cN.

But the form (3.2) is not the most well known form of the logistic equation. To obtain the most frequently used form, we transform (3.2) as follows:

$$\begin{aligned} N' &= (b-d)N - cN^2 \\ &= \big((b-d) - cN\big)N \\ &= \left(r - \frac{r}{r}cN\right)N, \quad \text{setting } r = b-d \\ &= rN\left(1 - \frac{c}{r}N\right) \\ &= rN\left(1 - \frac{N}{K}\right), \end{aligned}$$

with

$$\frac{c}{r} = \frac{1}{K}.$$

So, using the change of variables

$$(r, K) \leftrightarrow \left(b-d, \frac{b-d}{c}\right),$$

we transform (3.2) into the commonly used form

$$N' = rN\left(1 - \frac{N}{K}\right), \tag{3.3}$$

The parameter r is the **intrinsic growth rate**, K is the **carrying capacity**. There are three ways to tackle this equation:

i) The equation is separable. [explicit method]

ii) The equation is a Bernoulli equation. [explicit method]

iii) Use qualitative analysis.

3.3.1 Solving the logistic as a separable equation

3.3.2 Solving the equation as a Bernoulli equation

3.3.3 Qualitative analysis of the logistic equation

We study (3.3). For this, write

$$f(N) = rN\left(1 - \frac{N}{K}\right),$$

and consider the initial value problem (IVP)

$$\begin{aligned} N' &= f(N) \\ N(0) &= N_0 \geq 0. \end{aligned} \tag{3.4}$$

The function f is C^1 (differentiable with continuous derivative) so solutions to (3.4) exist and are unique, by virtue of Theorem **??**. **Equilibria** of (3.3) are points such that $f(N) = 0$ (so that $N' = f(N) = 0$, meaning N does not vary). So we solve $f(N) = 0$ for N. We find two points:

- $N = 0$,
- $N = K$.

We then consider the **well-posedness** of the problem, which consists in ensuring that solutions remain nonnegative (we are modelling populations) and bounded. This is usually carried out before any other type of analysis, but here, we use information derived from the nature of the equilibrium $N = 0$.

By uniqueness of solutions to (3.4), solutions cannot cross the curve $N(t) = 0$ (nor the curve $N(t) = K$, but this is not required for well-posedness). $N(t) = 0$ is a solution to (3.4), defined for all $t \geq 0$. Suppose that, for a solution through $N_0 > 0$, there exists $t = \tau > 0$ such that $N(\tau) = 0$, and that τ is the first value of t such that $N(t) = 0$ (recall that $N_0 > 0$). Then, at the point $(t, N) = (\tau, 0)$, we have two solutions: the solution through $(t, N) = (0, 0)$ and the solution through $(t, N) = (0, N_0)$. This is a contradiction, since solutions are known to be unique (and thus, through a given point (t, N), there is one, and one only, solution to (3.4)). Boundedness is easy to eastablish: remark that if $N > K$, then $f(N) < 0$, implying that solutions decrease for $N > K$.

For the general behavior of solutions, there are several cases to consider.

- If $N_0 = 0$, then $N(t) = 0$ for all $t \geq 0$, and from the above discussion, no solution with positive initial condition will ever reach zero.
- $N \in (0, K)$, then $rN > 0$ and $N/K < 1$ so $1 - N/K > 0$, which implies that $f(N) > 0$. As a consequence, $N(t)$ increases if $N_0 \in (0, K)$.
- If $N_0 = K$, then $N(t) = K$ for all $t \geq 0$.
- If $N > K$, then $rN > 0$ and $N/K > 1$, implying that $1 - N/K < 0$ and in turn, $f(N) < 0$. As a consequence, $N(t)$ decreases if $N_0 \in (K, +\infty)$.

Therefore, since the curve $N = K$ cannot be crossed,

Theorem 3.3.1. *Suppose that $N_0 > 0$. Then the solution $N(t)$ of (3.4) is such that*

$$\lim_{t\to\infty} N(t) = K,$$

so that K is the number of individuals that the environment can support, the ***carrying capacity*** *of the environment.*

If $N_0 = 0$, then $N(t) = 0$ for all $t \geq 0$.

3.4 The delayed logistic equation

Consider the logistic equation (3.2) written in the form

$$\frac{N'}{N} = (b-d) - cN.$$

Suppose that instead of instantaneous inhibition, there is a time delay τ between the instant the inhibiting event takes place and the moment when it affects the growth rate. For example, two individuals fight for food, and one later dies of the injuries sustained during this fight.

Reasonning as above, Hutchinson introduced in [7] a delayed logistic equation. In the case of a time τ between inhibiting event and inhibition, the equation above would be written as

$$\frac{N'}{N} = (b-d) - cN(t-\tau).$$

Using the change of variables introduced in the ordinary differential equation case, this is written

$$N'(t) = rN(t)\left(1 - \frac{N(t-\tau)}{K}\right). \tag{3.5}$$

Such an equation is called a **delay** differential equation. It is much more complicated to study than (3.3). In fact, although (3.3) and (3.5) look very similar and that (3.5) has been studied for about 60 years now, there are details about (3.5) that remain unknown to this day.

Delayed initial value problem The IVP takes the form

$$\begin{aligned} N'(t) &= rN(t)\left(1 - \frac{N(t-\tau)}{K}\right), \\ N(t) &= \phi(t) \text{ for } t \in [-\tau, 0], \end{aligned} \tag{3.6}$$

where $\phi(t)$ is some continuous function. Hence, initial conditions (called initial data in this case) must be specified on an interval, instead of being specified at a point, to guarantee existence and uniqueness of solutions.

We will not learn how to study this type of equation (this is graduate level mathematics). Some elementary notions are given in Chapter 13 and more detailed examples can be found in Chapter **??**.

To find equilibria, remark that delay should not play a role, since N should be constant. Thus, equilibria are found by considering the equation with no delay, which is (3.3).

Theorem 3.4.1. *Suppose that $r\tau < \pi/2$. Then solutions of (3.6) with positive initial data $\phi(t)$ starting close enough to K tend to K. If $r\tau < 37/24$, then all solutions of (3.6) with positive initial data $\phi(t)$ tend to K. If $r\tau > \pi/2$, then K is an unstable equilibrium and all solutions of (3.6) with positive initial data $\phi(t)$ on $[-\tau, 0]$ are oscillatory.*

There is a gray zone between $37/24$ ($\simeq 1.5417$) and $\pi/2$ ($\simeq 1.5708$). The global aspect was proved for $r\tau < 37/24$ in 1945 by Wright [18] using brute force computations. Although there is very strong numerical evidence that this is in fact true up to $\pi/2$, nobody has yet managed to prove it.

3.5 The logistic map

So far, we have seen continuous-time models, where $t \in \mathbb{R}$ (usually, $t \in \mathbb{R}_+$). Another way to model natural phenomena is by using a discrete-time formalism, that is, to consider equations of the form

$$x_{t+1} = f(x_t),$$

where $t \in \mathbb{N}$ or $\mathbb{Z}$, that is, t takes values in a discrete valued (countable) set. Time could for example be days, years, etc. This is called a **discrete-time system**, or a **difference equation**. Some notions of theory for difference equations are given in Chapter 5.

The logistic **map** is, for $t \geq 0$,

$$N_{t+1} = rN_t \left(1 - \frac{N_t}{K}\right). \tag{3.7}$$

To transform this into an initial value problem, we need to provide an initial condition $N_0 \geq 0$ at $t = 0$.

To derive equation (3.7), start with the ODE equation (3.3), but use the fact the the left hand side, dN/dt, can be represented by

$$\frac{dN}{dt} = \frac{N(t + \Delta t) - N(t)}{\Delta t},$$

with $\Delta t \to 0$. Let us normalize, and assume that the time step $\Delta t = 1$. Then, using (3.7),

$$N(t+1) - N(t) = rN(t)\left(1 - \frac{N(t)}{K}\right).$$

Rearranging,

$$N(t+1) = (r+1)N(t) - \frac{rN(t)^2}{K}.$$

Setting $\tilde{r} = r + 1$ and $\tilde{K} =$ and dropping the tildes, we obtain (3.7).

To study (3.7), we adimensionalize the model by using the change of variable

$$x_t = \frac{r}{K(1+r)}N(t)$$

to obtain, dropping the tilde,

$$x_{t+1} = (1+r)x_t(1 - x_t)$$

and $1 + r = \tilde{r}$.

For convenience we rewrite (3.7) as

$$x_{t+1} = rx_t(1 - x_t), \tag{3.8}$$

where r is a parameter in $\mathbb{R}_+$, and x is typically taken in $[0, 1]$. We let

$$f_r(x) = rx(1 - x). \tag{3.9}$$

This defines the discrete time logistic equation

$$x_{t+1} = f_r(x_t), \tag{3.10}$$

the latter being considered with initial condition $x_0 \in [0, 1]$.

3.5.1 Well-posedness

We consider the case $0 < r < 4$, where we know for certain that the iterates of f_r remain in the set $[0, 1]$. Indeed,

$$f_r'(x) = r - 2rx = r(1 - 2x), \tag{3.11}$$

so f_r is increasing for $x < 1/2$ and decreasing for $x > 1/2$, with a maximum at $x = 1/2$, equal to $r/4$. On the other hand, $f_r(0) = f_r(1) = 0$, so the minima are at $x = 0$ and $x = 1$. Therefore, if $r \leq 4$ then $f_r([0, 1]) \subseteq [0, 1]$. However, there are a few cases that can be excluded.

- If $x_0 = 0$, then $x_t = 0$, for all $t \geq 1$ and all r.
- If $x_0 = 1$, then $x_1 = 0$, and thus $x_t = 0$ for all $t \geq 1$ and all r. This is true for all t and all r: if there exists t_k such that $x_{t_k} = 1$, then $x_t = 0$ for all $t \geq t_k$.
- In the case $r = 4$, this last situation occurs if $x_t = 1/2$ for some t.
- Finally, if $r = 0$, then $x_t = 0$ for all t.

For these reasons, we generally consider $x \in (0, 1)$ and $r \in (0, 4)$.

3.5.2 Fixed points of f_r

Fixed points of (3.9) are found by solving the fixed point equation

$$x = f_r(x).$$

The reasoning is similar to what is done with ordinary differential equations. A solution remains fixed if $x_{t+1} = x_t$, which, when using $x_{t+1} = f_r(x_t)$, means that we must have $x_t = f_r(x_t)$, or, in other words, $x = f_r(x)$.

The fixed point equation is here

$$x = rx(1 - x).$$

It is clear that there are two points that satisfy this equation, namely $x = 0$ and $x = (r - 1)/r$. We denote from now on $p = (r - 1)/r$. Note that $x = 0$ always exists. On the other hand, p has the following properties:

- $\lim_{r\to 0^+} p = -\infty$.
- $\dfrac{\partial}{\partial r}p = \dfrac{1}{r^2} > 0$, so p is an increasing function of r.
- $p = 0$ if and only if $r = 1$ (unique since p is increasing).
- $\lim_{r\to\infty} p = 1$.

Remember that we are modelling a population, so we want $p > 0$ (or at least, nonnegative). If $p > 0$, we say that p is *biologically relevant*. For this, we need $r > 1$. In the case that $r < 1$, then p does exist, but we do not consider it, as it is not biologically relevant, and by abuse of language, say that p does not exist.

Conclusion 1. At this point, the situation is as follows. The fixed point $x = 0$ always exists, and

- if $r \in (0, 1)$, then p does not exist,
- if $r > 1$, then p exists.

3.5.3 Stability of the fixed points

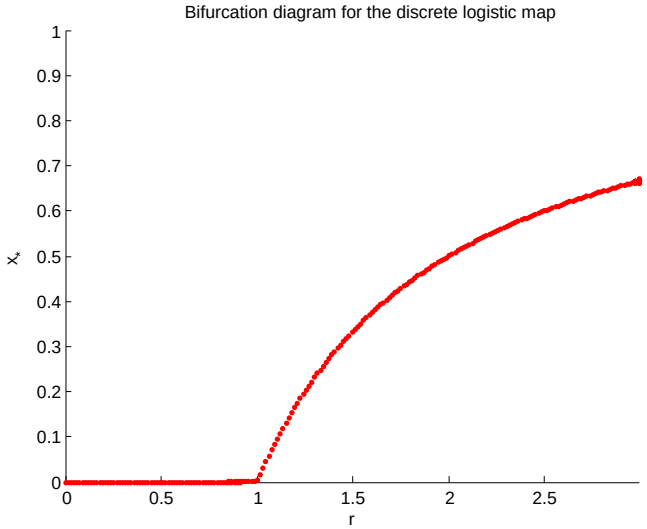
To determine the stability of f_r at a fixed point x^*, we need to compare $|f_r'(x^*)|$ with the value 1. From (3.11),

$$|f_r'(0)| = |r| = r,$$

and

$$\begin{aligned} |f_r'(p)| &= \left| r - 2r\frac{r-1}{r} \right| \\ &= |r - 2(r-1)| \\ &= |2 - r|. \end{aligned}$$

As a consequence, $x = 0$ is attracting if $r < 1$ and repelling otherwise, and $p = (r-1)/r$ is attracting if $|2 - r| < 1$, that is, $r < 3$, and repelling otherwise.



Conclusion 2. Building upon **Conclusion 1**, we therefore deduce that

- if $r \in (0,1)$, then $x = 0$ is attracting, and the fixed point $x = p$ does not exist,
- if $r \in (1,3)$, then $x = 0$ is repelling, and the fixed point $x = p$ exists and is attracting,
- if $r > 3$, then $x = 0$ is repelling, and the fixed point $x = p$ exists and is repelling.

Remark – A fixed point that is such that $|f'(p)| \neq 1$, or a periodic point such that $|(f^k)'(p)| \neq 1$, is called *hyperbolic*. In the case that $|f'(p)| = 1$, or, for a periodic point, $|(f^k)'(p)| = 1$, then p is called *non hyperbolic*. The non hyperbolic case is harder to treat. Here, the case $r = 1$ is a *non hyperbolic* case. However, the probability that $r = 1$ is zero (the set $r = \{1\}$ has measure zero in the parameter domain $0 < r < 4$), explaining why, most of the time, the case of r taking a single value, such as $r = 1$, is omitted. ◦

3.5.4 Stable sets of the fixed points

Conclusion 2 establishes that $x = 0$ and $x = p$ are attracting when, respectively, $r \in (0,1)$ and $r \in (1,3)$. This is not sufficient to characterize the behavior of all solutions. Remember that attractiveness of a fixed point x^* implies that there is a neighborhood of x^* that belongs to $W^s(x^*)$, i.e., there exists a neighborhood $\mathcal{N} \ni x^*$ such that $\forall x \in \mathcal{N}$, x is forward asymptotic to x^*.

If we want to make sure that we have the "complete picture", we need to show that $W^s(x^*) = [0,1]$, i.e., that all solutions go to x^*. There are several ways to tackle this problem, only one is shown here.

Case of the fixed point $x = 0$ (i.e., case $0 < r < 1$)

Since $r < 1$, it follows that $f_r(x) = rx(1-x) < x(1-x)$ Also, $x \in [0,1]$ implies that $1-x \in [0,1]$, and therefore $f_r(x) < x(1-x) \leq x$. Therefore, for any $x_0 \in [0,1]$,

$$\begin{aligned} x_1 &= f_r(x_0) \\ &< x_0 \\ x_2 &= f_r(x_1) \\ &< x_1. \end{aligned}$$

Therefore we have a strictly decreasing sequence. Since $[0,1]$ is invariant, the sequence is bounded below by 0. Therefore $\lim_{k\to\infty} f^k(x_0) = 0$, and $W^s(0) = [0,1]$ when $0 < r < 1$. Therefore we can strengthen **Conclusion 2**.

Conclusion 3′. If $0 < r < 1$, then for all $x_0 \in [0,1]$, $\lim_{k\to\infty} f^k(x_0) = 0$, or, equivalently, $\lim_{t\to\infty} x_t = 0$.

Remark – In the considerations above, we could have used Theorem 5.6.15 directly, once it was established that $f_r(x) < x$. ◦

Case of the fixed point $x = p$ (i.e., case $1 < r < 3$)

We know that f_r is increasing from a minimum of 0 at $x = 0$ to a maximum $r/4$ at $x = 1/2$, and then decreasing from there to a minimum of 0 at $x = 1$. Therefore, we must distinguish two cases: $r < 2$ and $r > 2$.

Case $1 < r < 2$ In the case $r < 2$, the intersection of $f_r(x)$ with the first bisectrix x occurs before the maximum $r/4$ is reached (see Figure 3.1). Then we are in a position to use part **a**

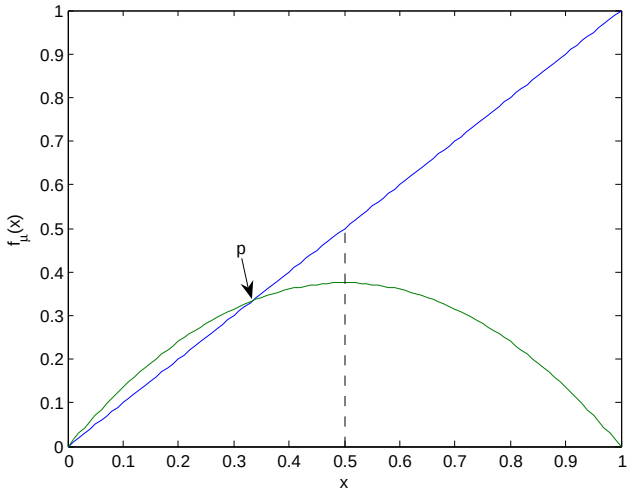


Figure 3.1: The function $f_{1.5}(x)$.

in Theorem 5.6.19, giving the conclusion.

Case $2 < r < 3$ In the case $r > 2$, the intersection of $f_r(x)$ with the first bisectrix x occurs after the maximum $r/4$ is reached (see Figure 3.2). We use Theorem 5.6.16 in Appendix 5.5.3.

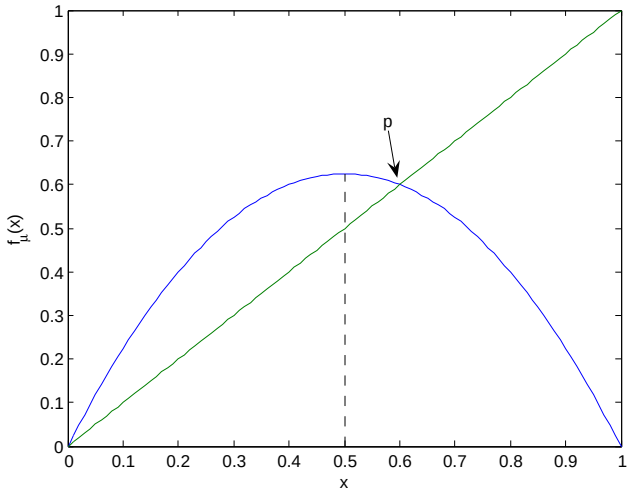


Figure 3.2: The function $f_{2.5}(x)$.

In order to be able to apply this theorem, we must first prove that there are no 2-cycles. For this purpose, we could try to use Theorem 5.6.17, but it fails to provide the conclusion here. Here, we can use the result from Section 3.5.5, where it is shown by explicit calculation that there are no 2-cycles for $r < 3$. It is clear that f satisfies the hypotheses of Theorem 5.6.16. Indeed, f is decreasing for $x > 1/2$.

Remark – We could also have used part **b** in Theorem 5.6.19, but this required to show that $f^2(x) > x$ for $x \in (1/2, p)$. ∘

Global stability By theorem 5.6.19 the equilibrium p is globally asymptotically stable for $1 < \mu < 3$.

Case $r = 2$ The result used to treat the case $r < 2$ can, in the present case, be extended to include the case where $r = 2$.

3.5.5 Existence of points of period 2

We now study the existence of periodic points with least period 2, that is, fixed points of the map $f_r^2(x)$. We have

$$\begin{aligned} f_r^2(x) &= f_r(f_r(x)) \\ &= r f_r(x)(1 - f_r(x)) \\ &= r^2 x(1-x)(1 - rx(1-x)). \end{aligned} \tag{3.12}$$

Remark that 0 and p are points of period 2. Indeed, a fixed point x^* of f satisfies $f(x^*) = x^*$, and as a consequence, $f^2(x^*) = f(f(x^*)) = f(x^*) = x^*$. This is extremely helpful in localizing the other periodic points, if there are any. Indeed, writing the fixed point equation as

$$f_r^2(x) - x = 0,$$

and defining $Q(x) := f_r^2(x) - x$, we see that, since 0 and p are fixed points of f_r^2, they are roots of $Q(x)$. Therefore, Q can be factorized as

$$Q(x) = x(x-p)(-r^3x^2 + Bx + C),$$

since it is clear from (3.12) that f_r^2 is a polynomial of degree 4 with leading coefficient equal to $-r^3$. Substituting the value $(r-1)/r$ for p in Q, developing Q and (3.12) and equating coefficients of like powers gives

$$Q(x) = x\left(x - \frac{r-1}{r}\right)\left(-r^3x^2 + r^2(r+1)x - r(r+1)\right). \tag{3.13}$$

The roots of (3.13) are the fixed points of f_r^2. Since $x = 0$ and $x = p$ are already known, we can concentrate on the roots of the polynomial

$$R(x) := -r^3x^2 + r^2(r+1)x - r(r+1).$$

The discriminant is $\Delta = r^4(r+1)^2 - 4r^4(r+1) = r^4(r+1)(r+1-4) = r^4(r+1)(r-3)$. Therefore, R has no real root if $r < 3$, a double real root if $r = 3$ and distinct real roots if $r > 3$. We want real valued solutions, so discard the case $r < 3$. In the case $r = 3$, the root is

$$\left.\frac{r+1}{2r}\right|_{r=3} = \frac{2}{3},$$

which is the value of p when evaluated at $r = 3$: the fixed point p and the fixed point deduced from R coincide at $r = 3$.

So we now consider the case $r > 3$. In this case, R has two distinct real roots (that is, f_r^2 has two distinct real fixed points) given by

$$\bar{x}_{1,2} = \frac{r+1 \pm \sqrt{(r+1)(r-3)}}{2r}.$$

First, remark that for $r > 3$ but very close to 3, it follows from the continuity of R that the roots are very close to the double root $2/3$, and hence are in $(0,1)$. More than the actual value of the roots, what is of interest at this point is to determine whether they remain in $(0,1)$ for all values of $3 < r < 4$. If a root is not in $(0,1)$, it is considered to be non biologically relevant and therefore is ignored.

To show that the roots remain in $(0,1)$ as we move away from 3, we could proceed directly using the expression for $\bar{x}_{1,2}$. Instead, we use Descartes' rule of signs (Theorem A.1, Appendix A). For this, remark that R has signed coefficients $-+-$, giving two sign changes and the possibility of 0 or 2 positive real roots. On the other hand, $R(-x)$ has signed coefficients $---$, hence there are no negative real roots. As we are in the case where the roots are real, it follows that both roots are positive.

To show that the roots are also smaller than 1, consider the change of variables $z = x - 1$. The polynomial R is transformed into

$$\begin{aligned} R_2(z) &= -r^3(z+1)^2 + r^2(r+1)(z+1) - r(r+1) \\ &= -r^3 z^2 + r^2(1-r)z - r. \end{aligned}$$

For $r > 1$, all the coefficients of this polynomial are negative, implying that R_2 has no root $z > 0$, implying in turn that R has no root $x > 1$.

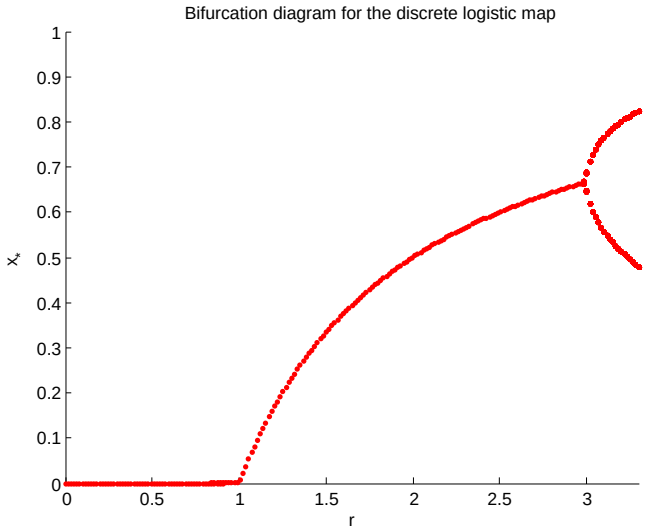


Figure 3.3: Bifurcation diagram for the discrete logistic map showing the birth of the fixed point at $r = 1$ and of the 2-cycle at $r = 3$.

3.5.6 Attractiveness of the periodic orbit

We use Theorem **??** which states that the $2-$cycle is locally asymptotically stable if

$$|f_r'(\bar{x}_1) f_r'(\bar{x}_2)| < 1.$$

From (3.11), we obtain

$$\begin{aligned}|f_r'(\bar{x}_1)f_r'(\bar{x}_2)| &= \left|\left(r-(r+1)+\sqrt{(r+1)(r-3)}\right)\left(r-(r+1)-\sqrt{(r+1)(r-3)}\right)\right| \\ &= \left|\left(1+\sqrt{(r+1)(r-3)}\right)\left(1-\sqrt{(r+1)(r-3)}\right)\right|.\end{aligned}$$

Therefore,

$$\begin{aligned}|f_r'(\bar{x}_1)f_r'(\bar{x}_2)| < 1 &\Leftrightarrow |1-(r+1)(r-3)| < 1 \\ &\Leftrightarrow -1 < 1-(r+1)(r-3) < 1 \\ &\Leftrightarrow 0 < (r+1)(r-3) < 2.\end{aligned}$$

Evidently, the inequality $0 < (r+1)(r-3)$ is satisfied if $r > 3$. On the other hand, the quadratic inequality $(r+1)(r-3) < 2$ is satisfied if $r < 1+\sqrt{6}$. Therefore, the 2-cycle is locally asymptotically stable if

$$3 < r < 1+\sqrt{6}$$

and unstable if $r > 1+\sqrt{6}$.

3.5.7 The cascade of bifurcation to chaos

We have seen that at $r = 1$, $r = 3$ and $r = 1+\sqrt{6}$, there are changes in the stability of the various equilibria that are present, and that additional equilibria can be born. These values of r are called **bifurcation points**.

The first bifurcation that occurs, at $r = 1$, is different in nature from the ones that follow. It is called a **transcritical bifurcation**, and corresponds to an exchange of stability between zero and p (recall that although p is not relevant for $r < 1$, it still does exist).

Subsequent bifurcations are called **period-doubling bifurcations**. By continuing the analysis of the logistic map, we see that it undergoes a a sequence of period doubling bifurcations, called the **period-doubling cascade**, as r increases from 3 to 4 (see Figure 3.4):

- for $1+\sqrt{6} < r < 3.5441$ there is a stable $4-$cycle, followed by a period doubling at $r = 3.5441$;
- for $3.5441 < r < 3.5644$ there is a stable $8-$cycle, followed by a period doubling at $r = 3.5644$;
- for $3.5644 < r < 3.5688$ there is a stable $16-$cycle, followed by a period doubling at $r = 3.5688$;
- ... other stable cycles of increasing period 2^n;
- finally, for $r > 3.57$, a cycle of period 3 exists. In that case, the solutions are called **chaotic** (see below).

The points at which the period doublings occur has a very interesting (and intriguing) property: The bifurcation points form a sequence, $\{r_n\}$, that has the property that

$$\lim_{n\to\infty}\frac{r_n - r_{n-1}}{r_{n+1}-r_n}$$

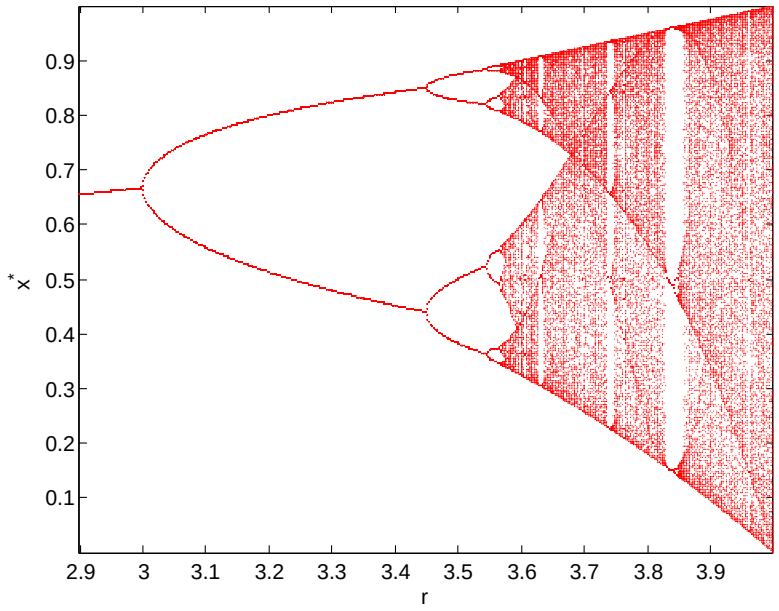


Figure 3.4: Bifurcation cascade for $2.9 \leq r \leq 4$.

exists and is a constant, called the Feigenbaum constant, equal to 4.669202... This constant has been shown to be the same in many of the maps that undergo the same type of cascade of period doubling bifurcations.

To finish, let us briefly discuss chaos. Note that the mathematics are quite involved and well beyond the scope of these notes. Denoting $\triangleright$ an order symbol, Sharkovskii's ordering of integers is as follows:

$$3 \triangleright 5 \triangleright 7 \triangleright 9 \triangleright 11 \triangleright \cdots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright \cdots \triangleright 2 \cdot 9 \triangleright \cdots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright \cdots$$
$$\triangleright 2^n \cdot 3 \triangleright 2^n \cdot 5 \triangleright \cdots \triangleright 2^{n+1} \cdot 3 \triangleright 2^{n+1} \cdot 5 \triangleright \cdots \triangleright 2^{n+1} \triangleright 2^n \triangleright \cdots 2^2 \triangleright 2 \triangleright 1.$$

This gives an ordering of all positive integers. The following result helps characterizing the behavior of the cascade.

Theorem 3.5.1 (Sharkovskii)**.** *Let $f : I \subset \mathbb{R} \to \mathbb{R}$ be a continuous function. Assume that f has a point of least period n and $n \triangleright k$. Then f has a point of least period k.*

We have seen that for $r > 3.57$, there is a cycle of period 3 for the logistic map. By Sarkovskii's theorem, the presence of period 3 points implies the presence of points of all periods. At this point, the system is said to be in a **chaotic regime**, or **chaotic**.

3.6 Conclusion, extensions and further reading

We have used three different modelling paradigms to describe the growth of a population in a **logistic** framework:

- The ODE version of Section 3.3 has monotone solutions converging to the carrying capacity K.

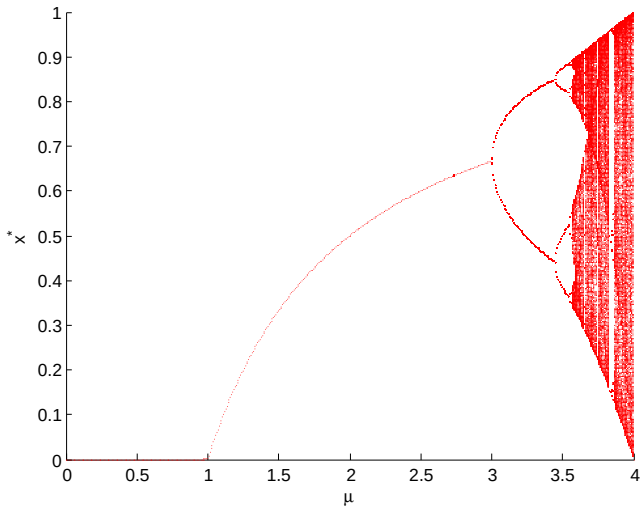


Figure 3.5: The cascade of bifurcation to chaos for the logistic map, for r varying between 0 and 4.

- The DDE version of Section 3.4 has oscillatory solutions, either converging to K or, if the delay is too large, periodic about K.
- The discrete time version of Section 3.5 has all sorts of behaviors, and can be chaotic.

It is important to be aware that the **choice of modelling method** is almost **as important** in the outcome of the model as the precise formulation/hypotheses of the **model**.

Chapter 4

Time of residence in states

Introduction to compartmental models

4.1 Time spent in a state – Some probability theory

We suppose that a system/object/individual can be in two states S_1 and S_2. These states can be anything:

- S_1: working, S_2: broken,
- S_1: infected, S_2: recovered,
- S_1: alive, S_2: dead.

At time $t = 0$, the system is in state S_1. An event happens at some time $t = \tau \geq 0$, which triggers the switch from state S_1 to state S_2. Suppose that we are able to conduct an experiment with infinitely many copies of this system, and are interested in obtaining some type of description of the time it takes for the system to switch states.

Let us reformulate this problem in the language of probability theory. A **random variable** is a variable that takes random values, that is, a mapping from random experiments to numbers. Here, let us call T the random variable "time spent in state S_1 before switching into state S_2". We take a collection of objects or individuals that are in state S_1 and want some law for the **distribution** of the times spent in S_1, i.e., a law for T. For example, we make light bulbs and would like to tell our customers that on average, our light bulbs last 200 years. For this, we conduct an **infinite** number of experiments, and observe the time that it takes, in every experiment, to switch between S_1 and S_2. From this, we deduce a model, which in this context is called a **probability distribution**.

Before we proceed any further, let us make one remark here. Roughly speaking, probability theory assumes that it is indeed possible to carry out an infinite number of experiments. Statistics, on the other hand, deals with the use of probability theory in "real life", where the number of experiments is intrinsically limited.

Returning to the time spent in a state, we assume that T is a **continuous** random variable, that is, T takes continuous values. Examples of continuous random variables are the height or age of a person (if measured very precisely), distances, times, etc. Another type of random variables are **discrete** random variables, which take values in a denumerable set: heads or tails

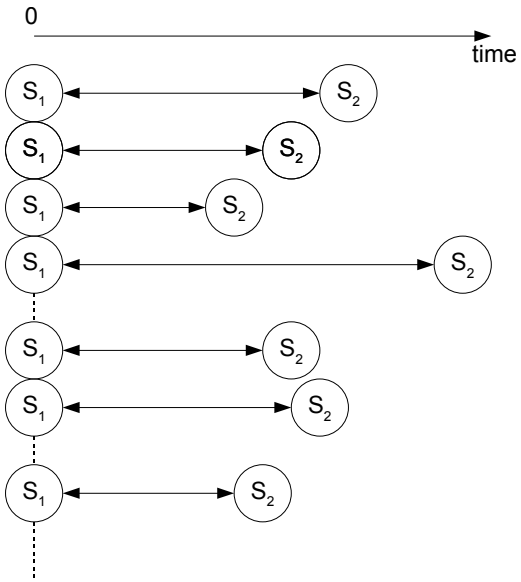


Figure 4.1: Time until the switch from state S_1 to state S_2, for several experiments. From this, we could deduce a distribution for the time of sojourn in state S_1.

on a coin toss, the number rolled on a dice, the height of a person, if expressed rounded without subunits, the age of a person in years (without subunits), etc.

A **probability** is a function $\mathcal{P}$, from a probability space to $[0, 1]$. Formally: $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space, with Ω the **sample** space, $\mathcal{F}$ a σ-algebra of subsets of Ω whose elements are the **events**, and $\mathcal{P}$ a **measure** from $\mathcal{F}$ to $[0, 1]$ such that $\mathcal{P}(E) \geq 0$, $\forall E \subset \Omega$, $\mathcal{P}(\Omega) = 1$ and $\mathcal{P}(E_1 \cup E_2 \cup \cdots) = \sum_i \mathcal{P}(E_i)$.

A probability gives the likelihood of an event occurring, among all the events that are possible, in that particular setting. For example,

$$\mathcal{P}(\text{getting heads when tossing a coin}) = 1/2$$

and

$$\mathcal{P}(\text{getting tails when tossing a coin}) = 1/2.$$

Since T is continuous, it has a continuous **probability density function** (p.d.f.), f, which satisfies:

- $f \geq 0$,
- $\int_{-\infty}^{+\infty} f(s)ds = 1$.
- $\mathcal{P}(a \leq T \leq b) = \int_a^b f(t)dt$.

The **cumulative distribution function** (c.d.f.) is a function $F(t)$ that characterizes the distribution of T, and defined by

$$F(s) = \mathcal{P}(T \leq s) = \int_{-\infty}^{s} f(x)dx.$$

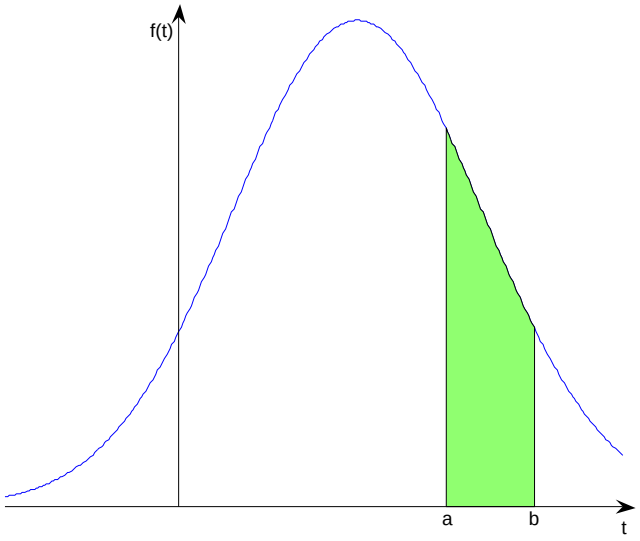


Figure 4.2: A continuous probability distribution function (curve) and $\mathcal{P}(a \leq T \leq b)$ (area).

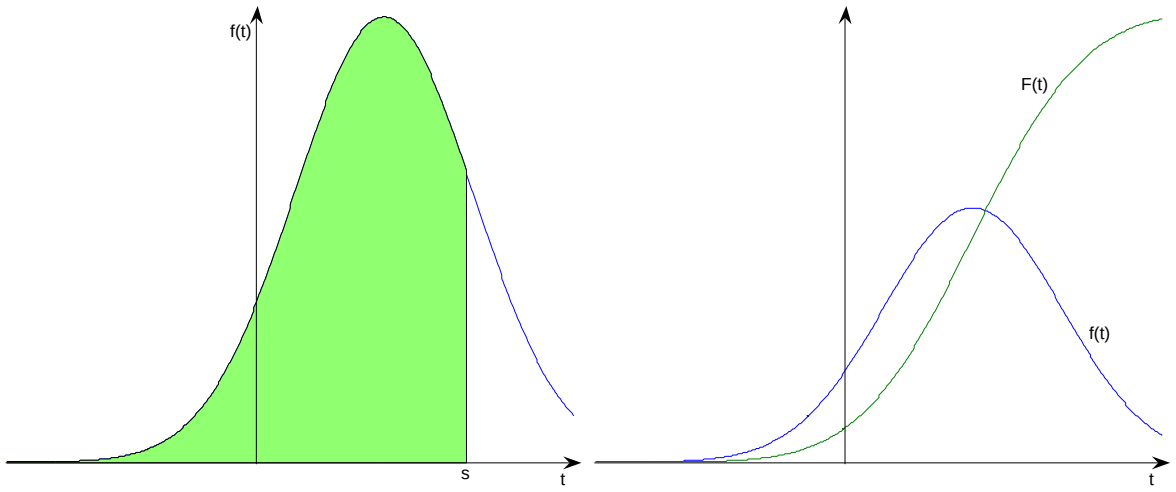


Figure 4.3: The link between the cumulative distribution function and the probability density function. The integral from $-\infty$ to t, shown on the left by the area under the curve f, gives a function $F(t)$ (right).

- Since f is a nonnegative function, F is nondecreasing.
- Since f is a probability density function, $\int_{-\infty}^{+\infty} f(s)ds = 1$, and thus $\lim_{t\to\infty} F(t) = 1$.

For a continuous random variable T with probability density function f, the **mean** value of T, denoted $\bar{T}$ or $E(T)$, is given by

$$E(T) = \int_{-\infty}^{+\infty} tf(t)dt.$$

Another characterization of the distribution of the random variable T is through the **survival** (or **sojourn**) function. The survival function of state S_1 is given by

$$\mathcal{S}(t) = 1 - F(t) = \mathcal{P}(T > t) \tag{4.1}$$

This gives a description of the **sojourn time** of a system in a particular state (the time spent in the state). $\mathcal{S}$ is a nonincreasing function (since $\mathcal{S} = 1 - F$ with F a c.d.f.), and $\mathcal{S}(0) = 1$ (since T is a positive random variable). The **average sojourn time** τ in state S_1 is given by

$$\tau = E(T) = \int_0^{\infty} tf(t)dt$$

Assuming that $\lim_{t\to\infty} t\mathcal{S}(t) = 0$ (which is verified for most probability distributions),

$$\tau = \int_0^{\infty} \mathcal{S}(t)dt.$$

4.2 The exponential distribution

The random variable T has an **exponential** distribution if its probability density function takes the form

$$f(t) = \begin{cases} 0 & \text{if } t < 0, \\ \theta e^{-\theta t} & \text{if } t \geq 0, \end{cases} \tag{4.2}$$

with $\theta > 0$. Then the survival function for state S_1 is of the form $\mathcal{S}(t) = e^{-\theta t}$, for $t \geq 0$, and the average sojourn time in state S_1 is

$$\tau = \int_0^{\infty} e^{-\theta t}dt = \frac{1}{\theta}$$

If on the other hand, for some constant $\omega > 0$,

$$\mathcal{S}(t) = \begin{cases} 1, & 0 \leq t \leq \omega \\ 0, & \omega < t \end{cases}$$

which means that T has a Dirac delta distribution $\delta_\omega(t)$, then the average sojourn time is a constant, namely

$$\tau = \int_0^{\omega} dt = \omega$$

These two distributions can be regarded as extremes.

4.3 A cohort model

We consider a population consisting of individuals born at the same time (a **cohort**), for example, the same year. We suppose

- At time $t = 0$, there are initially $N_0 > 0$ individuals.
- All causes of death are compounded together.
- The time until death, for a given individual, is a random variable T, with continuous probability density distribution $f(t)$ and survival function $P(t)$.

Denote $N(t)$ the population at time $t \geq 0$. Then

$$N(t) = N_0 P(t). \tag{4.3}$$

- $N_0 P(t)$ gives the proportion of N_0, the initial population, that is still alive at time t.

Case where T is exponentially distributed Suppose that T has an exponential distribution with mean $1/d$ (or parameter d), $f(t) = de^{-dt}$. Then the survival function is $P(t) = e^{-dt}$, and (4.3) takes the form

$$N(t) = N_0 e^{-dt}. \tag{4.4}$$

Now note that

$$\begin{aligned} \frac{d}{dt}N(t) &= -dN_0 e^{-dt} \\ &= -dN(t), \end{aligned}$$

with $N(0) = N_0$. Thus, the ODE $N' = -dN$ makes the assumption that the life expectancy at birth is exponentially distributed.

Case where T has a Dirac delta distribution Suppose that T has a Dirac delta distribution at $t = \omega$, giving the survival function

$$P(t) = \begin{cases} 1, & 0 \leq t \leq \omega, \\ 0, & t > \omega. \end{cases}$$

Then (4.3) takes the form

$$N(t) = \begin{cases} N_0, & 0 \leq t \leq \omega, \\ 0, & t > \omega. \end{cases} \tag{4.5}$$

All individuals survive until time ω, then they all die at time ω. Here, we have $N' = 0$ everywhere except at $t = \omega$, where it is undefined.

4.4 Sojourn times in an SIS disease transmission model

The population can be

- closed (no immigration, no emigration, birth and death are neglected).
- open
- homogeneous and homogeneously mixed (or not)

The population has a structure

- classification of individuals according to their disease status
  - susceptible (S): individuals not infective but who are capable of contracting the disease
  - latent or exposed (E): infected by the disease, but not yet infectious
  - infective (I): infectious individual; an individual can be infectious before symptoms appear.
  - removed (R): no longer infectious, whether by acquiring immunity or death...
  - carrier: in some diseases, individual can remain infectious for long periods (e.g. for life), but do not show any symptoms of the disease themselves.
- age
- sex

Types of models

- SI model: no recovery.
- SIS model: recovery but no immunity.
- SIR model: recovery with permanent immunity.
- SIRS model: recovery with temporary immunity.
- ...

Consider a disease and a a population of individuals who can be infected by this disease. Both can be anything: a human population subject to influenza, an animal population subject to foot and mouth disease, a rumor spreading in a human population, inovation spreading through businesses, a computer virus spreading on the internet..

Suppose that individuals can be identified with respect to their epidemiological status:

- susceptible to the disease,
- infected by the disease,
- recovered from the disease.

These states are clearly of the type we were discussing in Section 4.1. To be more specific, consider a disease that confers no immunity. In this case, individuals are either

- **susceptible** to the disease, with the number of such individuals at time t denoted by $S(t)$,
- or **infected** by the disease (and are also **infective** in the sense that they propagate the disease), with the number of such individuals at time t denoted by $I(t)$.

We want to model the evolution with time of S and I (t is omitted unless necessary). **Extremely important:** State all your hypotheses.

Hypotheses

- Individuals typically recover from the disease.
- The disease does not confer immunity.
- There is no birth or death.
- Infection is of **standard incidence** type

Once your hypotheses are stated, detail them if need be.

Recovery and No immunity

Individuals recover from the disease: the infection is not permanent. Upon recovery from the disease, an individual becomes susceptible again immediately. Good description for diseases that confer no immunity, e.g., the cold or gonorrhea.

No birth or death Suppose that

- the time period of interest is short,
- the population is large enough,

then it is reasonable to assume that the total population is constant, in the absence of disease.

For mild diseases (cold, etc.), there are very little risks of dying from the disease. We assume no disease-induced death.

Hence $N \equiv N(t) = S(t) + I(t)$ is the (constant) total population.

Standard incidence New infectives result from random contacts between susceptible and infective individuals, described using standard incidence:

$$\beta \frac{SI}{N},$$

- $\beta SI/N$ is a rate (per unit time),
- β is the **transmission coefficient**, giving probability of transmission of the disease in case of a contact, times the number of such contacts made by an infective per unit time.

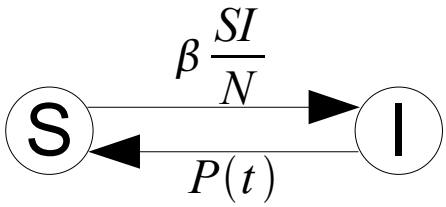


Figure 4.4: The flow diagram of the model with arbitrary infectious period.

Recovery We have not yet stated our hypotheses on the recovery process.. Traditional epidemiological models assume recovery from disease with a rate constant γ. Here, assume that, of the individuals who have become infective at time t_0, a fraction $P(t - t_0)$ remain infective at time $t \geq t_0$. Thus, considered for $t \geq 0$, the function $P(t)$ is a survival function.

Figure 4.4 is a **flow diagram** of our model. It details the flows of individuals between the compartments in the system. It is extremely useful to rapidly understand what processes are modelled.

Reducing the dimension of the problem To formulate our model, we would in principle require an equation for S and an equation for I.

But we have

$$S(t) + I(t) = N, \text{ or equivalently, } S(t) = N - I(t).$$

N is constant (equal total population at time $t = 0$), so we can deduce the value of $S(t)$, once we know $I(t)$, from the equation $S(t) = N - I(t)$.

We only need to consider 1 equation. **Do this when possible!** (nonlinear systems are hard, one less equation can make a lot of difference)

Model for infectious individuals Integral equation for the number of infective individuals:

$$I(t) = I_0(t) + \int_0^t \beta \frac{(N - I(u))I(u)}{N} P(t - u)du \tag{4.6}$$

- $I_0(t)$ number of individuals who were infective at time $t = 0$ and still are at time t.
  - $I_0(t)$ is nonnegative, nonincreasing, and such that $\lim_{t\to\infty} I_0(t) = 0$.
- $P(t - u)$ proportion of individuals who became infective at time u and who still are at time t.
- $\beta(N - I(u))S(u)/N$ is $\beta S(u)I(u)/N$ with $S(u) = N - I(u)$, from the reduction of dimension.

Expression under the integral Integral equation for the number of infective individuals:

$$I(t) = I_0(t) + \int_0^t \beta \frac{(N - I(u))I(u)}{N} P(t - u)du \tag{4.6}$$

The term

$$\beta \frac{(N - I(u))I(u)}{N} P(t - u)$$

- $\beta(N - I(u))I(u)/N$ is the rate at which new infectives are created, at time u,
- multiplying by $P(t-u)$ gives the proportion of those who became infectives at time u and who still are at time t.

Summing over $[0, t]$ gives the number of infective individuals at time t.

Case of an exponentially distributed time to recovery Suppose that $P(t)$ is such that the sojourn time in the infective state has an exponential distribution with mean $1/\gamma$, *i.e.*, $P(t) = e^{-\gamma t}$.

Then the initial condition function $I_0(t)$ takes the form

$$I_0(t) = I_0(0)e^{-\gamma t},$$

with $I_0(0)$ the number of infective individuals at time $t = 0$. This is obtained by considering the cohort of initially infectious individuals, giving a model such as (4.3). Equation (4.6) becomes

$$I(t) = I_0(0)e^{-\gamma t} + \int_0^t \beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)} du. \tag{4.7}$$

Taking the time derivative of (4.7) yields

$$\begin{aligned} I'(t) &= -\gamma I_0(0)e^{-\gamma t} - \gamma \int_0^t \beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)} du + \beta \frac{(N - I(t))I(t)}{N} \\ &= -\gamma \left(I_0(0)e^{-\gamma t} + \int_0^t \beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)} du \right) + \beta \frac{(N - I(t))I(t)}{N} \\ &= \beta \frac{(N - I(t))I(t)}{N} - \gamma I(t), \end{aligned}$$

which is the classical logistic type ordinary differential equation (ODE) for I in an SIS model without vital dynamics (no birth or death).

Case of a step function survival function Consider case where the time spent infected has survival function

$$P(t) = \begin{cases} 1, & 0 \le t \le \omega, \\ 0, & t > \omega. \end{cases}$$

i.e., the sojourn time in the infective state is a constant $\omega > 0$.

In this case (4.6) becomes

$$I(t) = I_0(t) + \int_{t-\omega}^t \beta \frac{(N - I(u))I(u)}{N} du. \tag{4.8}$$

Here, it is more difficult to obtain an expression for $I_0(t)$. It is however assumed that $I_0(t)$ vanishes for $t > \omega$.

When differentiated, (4.8) gives, for $t \ge \omega$,

$$I'(t) = I_0'(t) + \beta \frac{(N - I(t))I(t)}{N} - \beta \frac{(N - I(t-\omega))\, I(t-\omega)}{N}.$$

Since $I_0(t)$ vanishes for $t > \omega$, this gives the delay differential equation (DDE)

$$I'(t) = \beta \frac{(N - I(t))I(t)}{N} - \beta \frac{(N - I(t-\omega))I(t-\omega)}{N}.$$

4.5 Conclusion

- The distribution of the time of sojourn in classes (compartments) plays an important role in determining the type of model that we deal with.
- All compartmental ODE models, when they use terms of the form κX, make the assumption that the time of sojourn in compartments is exponentially distributed.
- At the other end of the spectrum, delay differential equations with discrete delay make the assumption of a constant sojourn time, equal for all individuals.
- Both can be true sometimes.. but reality is often somewhere in between. See Figure 4.5 for an example of how an exponential distribution, with its fat tail, overepresents long sojourn times.

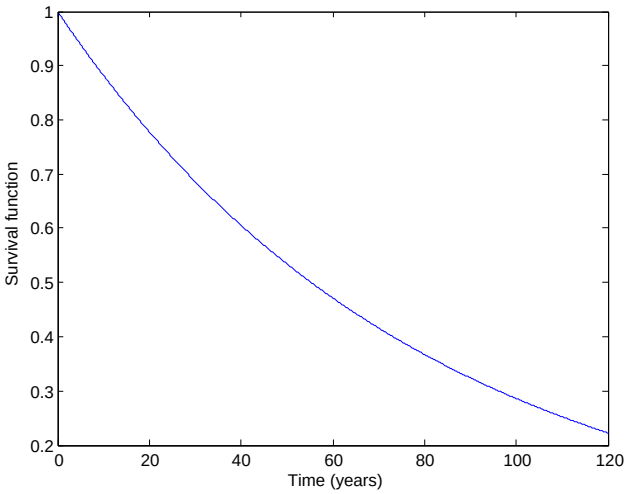


Figure 4.5: Survival function, $\mathcal{S}(t) = \mathcal{P}(T > t)$, for an exponential distribution with mean 80 years. Note that this implies that more than 20% of individuals in a cohort survive past the age of 120 years.

Part II

Deterministic discrete time systems

Chapter 5

A brief theory of discrete time systems

In this chapter, we consider discrete-time systems of the form

$$x_{t+1} = f(x_t), \tag{5.1}$$

with initial condition given for $t = 0$ by x_0, with $x, x_0 \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$. We also consider pth order equations of the form

$$f(x_{t+p}, x_{t+p-1}, \ldots, x_{t+1}, x_t, t) = 0, \tag{5.2}$$

where f is a real-valued function of the real variables x_t through x_{t+p} and t. We could also consider systems of nth order equations, but for the sake of brievety, we will limit ourselves to equations of the form (5.1) and (5.2).

Implicit in (5.1) and (5.2) is that the time interval is taken to be $\Delta t = 1$. Also, the state of the system at time t is denoted x_t. Formally, (5.1) should be written

$$x(t + \Delta t) = f(x(t)),$$

but this is cumbersome and will be used only when ambiguity could lead to miscomprehensions.

Using (5.1), we see that

$$\begin{aligned} x_1 &= f(x_0) \\ x_2 &= f(x_1) = f(f(x_0)) \stackrel{\Delta}{=} f^2(x_0) \\ &\vdots \\ x_k &= f^k(x_0). \end{aligned}$$

The compositions

$$f^k = \underbrace{f \circ f \circ \cdots \circ f}_{k \text{ times}}$$

are called the **iterates** of f. They define an infinite sequence of points

$$x_0, x_1, x_2, \ldots, x_t, \ldots,$$

that constitute the **solution** to (5.1). The object of this chapter is to determine the behavior of this solution. For example, in the case of the logistic map (3.9), do solutions behave like they do for the continuous time logistic equation (3.3) and tend to the carrying capacity K?

5.1 Types of equations/systems

Definition 5.1.1. *The* order *of a difference equation* (5.2) *is the difference between the largest and the smallest arguments p appearing in it.*

Remark that in biological terms, the order p of the equation is the number of previous generations that directly influence the value of x in a given generation.

Definition 5.1.2. *The difference equation is called* autonomous *if f does not depend explicitly on t and it is called* nonautonomous *otherwise.*

Definition 5.1.3. *Let*

$$x_{t+p} + a_1 x_{t+p-1} + a_2 x_{t+p-2} + \cdots + a_{p-1} x_{t+1} = b_t.$$

If the coefficients a_j, $j = 1, \ldots, p$ are constant or depend on t but ***do not*** *depend on the state variables, then the difference equation is said to be* linear*; otherwise, it is* nonlinear*.*

Definition 5.1.4. *If the difference equation is linear and $b_t = 0$ for all t, then it is said to be* homogeneous*; otherwise, it is said to be* nonhomogeneous*.*

Definition 5.1.5. *A solution of the difference equation*

$$f(x_{t+k}, x_{t+k-1}, \ldots, x_{t+1}, x_t, t) = 0$$

is a function x_t, $t = 0, 1, 2, \ldots$ such that when substituted into the equation makes it a true statement.

Some characteristics of difference equations

- changes of states are descibed over discrete intervals. Length of the discrete interval is some fixed length Δt: states of a system are modeled at the discrete time $t = 0, \Delta t, 2\Delta t, \ldots$
- recurrence relation
- evolutionary character or not
- to describe populations whose generations do not overlap:

5.2 First-order linear difference equation

Proposition 5.2.1. *Consider the first-order linear homogeneous difference equation with constant coefficients*

$$x_{t+1} = a x_t \tag{5.3}$$

If an initial value x_0 is known, the solution is unique and is given by

$$x_t = a^t x_0. \tag{5.4}$$

Proof. We have

$$x_1 = ax_0,$$

and so

$$x_2 = ax_1 = aax_0 = a^2x_0,$$

and

$$x_3 = ax_2 = aaax_0 = a^3x_0 \dots$$

Continuing, we have the general expression

$$x_t = a^t x_0.$$

□

Clearly, (5.4) defines a geometric sequence with common ratio a. Therefore, the **asymptotic behavior** of the solution depends on the value of a:

- if $|a| < 1$, then $\lim_{t\to\infty} x_t = 0$, i.e., x_t converges to 0,
- if $a = 1$, then for all $t \geq 0$, $x_t = x_0$, i.e., x_t remains constant,
- if $a = -1$, then for all $t \geq 0$, $x_t = (-1)^t x_0$, i.e., x_t alternates,
- if $|a| > 1$ then x_t diverges (either approaches infinity if $a > 1$ or diverges with alternating signs if $a < -1$).

Proposition 5.2.2. *Consider the first-order linear homogeneous difference equation defined for $t = 0, 1, 2, \dots$ by*

$$x_{t+1} = a_t x_t.$$

If an initial value x_0 is known, then the solution is unique and is given by

$$x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0.$$

Proof. Let us prove by mathematical induction that the proposition P_t holds $\forall t \in \mathbb{N} \setminus \{0\}$, with

$$P_t : \quad x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0.$$

First, we consider P_1. We have

$$x_1 = a_0 x_0,$$

hence P_1 is true. Then assume that P_t is true, i.e., $x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0$. Now express x_{t+1}:

$$\begin{aligned} x_{t+1} &= a_t x_t \\ &= a_t \left[\prod_{i=0}^{t-1} a_i\right] x_0 \quad \text{(by induction hypothesis)} \\ &= \left[\prod_{i=0}^{t} a_i\right] x_0, \end{aligned}$$

so P_{t+1} is true. By the principle of mathematical induction (PMI), we conclude that

$$x_t = \left[\prod_{i=0}^{t-1} a(i)\right] x_0, \forall t \in \mathbb{N} \setminus \{0\}.$$

□

Proposition 5.2.3. *Consider the first-order linear nonhomogeneous difference equation defined for $t = 0, 1, 2, \ldots$ by*

$$x_{t+1} = a_t x_t + b_t.$$

If an initial value x_0 is known, then the solution is unique and is given by

$$x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0 + b_{t-1} + \sum_{i=0}^{t-2} \left[\prod_{r=i+1}^{t-1} a_r\right] b_i.$$

In particular,

- *If $x_{t+1} = ax_t + b_t$, then*

$$x_t = a^t x_0 + \sum_{i=0}^{t-1} a^{t-i-1} b(i).$$

- *If $x_{t+1} = ax_t + b$, then*

$$x_t = \begin{cases} a^t x_0 + b\left[\dfrac{a^t - 1}{a - 1}\right] & a \neq 1 \\ x_0 + bt & a = 1. \end{cases}$$

Proof. Let us prove by mathematical induction that

$$P_t : \quad x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0 + b_{t-1} + \sum_{i=0}^{t-2} \left[\prod_{r=i+1}^{t-1} a_r\right] b_i$$

holds true for all $t \in \mathbb{N} \setminus \{0\}$. At rank $t = 2$: $x_1 = a_0 x_0 + b_0$, then

$$x_2 = a_1 x_1 + b_1 = a_1 a_0 x_0 + a_1 b_0 + b_1.$$

Now assume that P_t holds true, i.e.,

$$x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0 + b_{t-1} + \sum_{i=0}^{t-2} \left[\prod_{r=i+1}^{t-1} a_r\right] b_i$$

and express x_{t+1}:

$$\begin{aligned} x_{t+1} &= a_t x_t + b_t \\ &= a_t \left\{ \left[\prod_{i=0}^{t-1} a_i\right] x_0 + b_{t-1} + \sum_{i=0}^{t-2} \left[\prod_{r=i+1}^{t-1} a_r\right] b_i \right\} + b_t \\ &= \left[a_t \prod_{i=0}^{t-1} a_i\right] x_0 + a_t b_{t-1} + \sum_{i=0}^{t-2} \left[a_t \prod_{r=i+1}^{t-1} a_r\right] b_i + b_t \\ &= \left[\prod_{i=0}^{t} a_i\right] x_0 + b_t + a_t b_{t-1} + \sum_{i=0}^{t-2} \left[\prod_{r=i+1}^{t} a_r\right] b_i \\ &= \left[\prod_{i=0}^{t} a_i\right] x_0 + b_t + \sum_{i=0}^{t-1} \left[\prod_{r=i+1}^{t} a_r\right] b_i. \end{aligned}$$

Thus P_{t+1} holds. By the principle of mathematical induction, we conclude that, for all $t \in \mathbb{N} \setminus \{0\}$,

$$x_t = \left[\prod_{i=0}^{t-1} a_i\right] x_0 + b_{t-1} + \sum_{i=0}^{t-2} \left[\prod_{r=i+1}^{t-1} a_r\right] b_i. \quad \square$$

5.3 Higher-order linear equations

Definition 5.3.1. *The functions x_t^1, x_t^2,..., x_t^k are said to be linearly independent for $t \geq t_0$ whenever*

$$a_1 x_t^1 + a_2 x_t^2 + \cdots + a_k x_t^k = 0$$

for all $t \geq t_0$, then we must have $a_1 = a_2 = \cdots = a_k = 0$.

Definition 5.3.2. *The* Casoratian *of k functions x_t^1, x_t^2,..., x_t^k is defined as*

$$C(x_t^1, x_t^2, \ldots, x_t^k) = \det \begin{pmatrix} x_t^1 & x_t^2 & \ldots & x_t^k \\ x_{t+1}^1 & x_{t+1}^2 & \ldots & x_{t+1}^k \\ x_{t+2}^1 & x_{t+2}^2 & \ldots & x_{t+2}^k \\ \vdots & & & \\ x_{t+k-1}^1 & x_{t+k-1}^2 & \ldots & x_{t+k-1}^k \end{pmatrix}.$$

Using the Casoratian, it is easy to check the linear independence of solutions.

Proposition 5.3.3. *If the Casoratian of x_t^1, x_t^2,..., x_t^k satifies*

$$C(x_t^1, x_t^2, \ldots, x_t^k) \neq 0, \quad \forall t,$$

then x_t^1, x_t^2,..., x_t^k are k linearly independent functions.

Remark that the Casoratian and its implications are very similar to the Wronskian for the solutions of ordinary differential equations (Section **??**).

Definition 5.3.4. *A set of k linearly independent solutions of a k^{th}-order linear homogeneous difference equation is called a* fundamental set of solutions*.*

Proposition 5.3.5. *(Principle of superposition) If x_t^1, x_t^2,..., x_t^k are solutions of a k^{th}-order linear homogeneous difference equation, then*

$$c_1 x_t^1 + c_2 x_t^2 + \ldots, c_k x_t^k$$

is also solution of the k^{th}-order linear homogeneous difference equation.

Note that the principle of superposition also holds for ordinary differential equations (Section **??**).

Definition 5.3.6. *Let $\{x_t^1, x_t^2, \ldots, x_t^k\}$ be a fundamental set of solutions of k^{th} linear homogeneous difference equation. Then the general solution of the k^{th} linear homogeneous difference equation is given by*

$$x_t = \sum_{i=1}^{k} c_i x_t^i,$$

for arbitrary constants c_i, $i = 1, \ldots, k$

5.3.1 Homogeneous equations with constant coefficients

The case of a second-order equation is treated in detail to illustrate this subsection. A second-order linear homogeneous equation with constant coefficients takes the form

$$a_0 x_{t+2} + a_1 x_{t+1} + a_2 x_t = 0 \tag{5.5}$$

To find two linearly independent solutions, x_t^1 and x_t^2, assume that solutions take the form of $x_t = \lambda^t$, with $\lambda \neq 0$. Then substitute this tentative solution (*ansatz*) in (5.5),

$$a_0 \lambda^{t+2} + a_1 \lambda^{t+1} + a_2 \lambda^t = 0,$$

that is, since $\lambda^t \neq 0$,

$$a_0 \lambda^2 + a_1 \lambda + a_2 = 0.$$

The equation $a_0\lambda^2 + a_1\lambda + a_2 = 0$ is the **characteristic equation** of (5.5). The 2 roots of the characteristic equation, λ_1 and λ_2, are called the **eigenvalues**.

The general solution is a linear combination of the 2 solutions $x_t^1 = \lambda_1^t$ and $x_t^2 = \lambda_2^t$. The form of the general solution depends on the eigenvalues; there are 3 cases:

The eigenvalues are real and distinct: $\lambda_1 \neq \lambda_2$. The general solution is

$$x_t = c_1 \lambda_1^t + c_2 \lambda_2^t,$$

with c_1 and c_2 arbitrary constants.

The eigenvalues are real and equal: $\lambda_1 = \lambda_2$. Then the 2 linearly independent solutions are $x_t^1 = \lambda_1^t$ $x_t^2 = t\lambda_2^t$. The general solution is

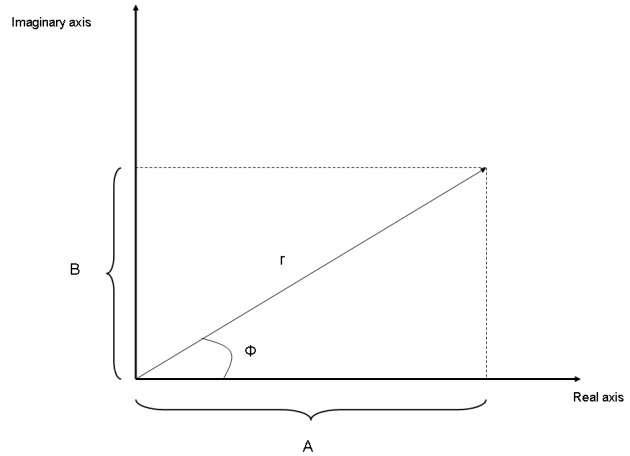
$$x_t = c_1 \lambda_1^t + c_2 t \lambda_2^t,$$

with c_1 and c_2 arbitrary constants.

The eigenvalues are complex conjugates: $\lambda_{1,2} = A \pm iB = r(\cos\phi \pm i \sin\phi)$, where $r = \sqrt{A^2 + B^2}$ and $\phi = \arctan(B/A)$. The two linearly independent solutions are $x^1 = r^t \cos(t\phi)$ and $x^2 = r^t \sin(t\phi)$. Then the general solution is

$$x_t = c_1 r^t \cos(t\phi) + c_2 r^t \sin(t\phi),$$

with c_1 and c_2 arbitrary constants.



An $m^{th}-$order linear homogeneous equation with constant coefficients is defined as

$$a_0 x_{t+m} + a_1 x_{t+m-1} + \cdots + a_m x_t = 0. \tag{5.6}$$

Solutions are composed of linear superpositions of m solutions of the form $x_t = \lambda^t$, $\lambda \neq 0$ where λ are obtained by finding the roots (eigenvalues) of the characteristic equation

$$a_0 \lambda^m + a_1 \lambda^{m-1} + \cdots + a_m = 0.$$

The characteristic equation has m eigenvalues: λ_i, $i = 1, \ldots, m$.

If eigenvalues are all real and distinct, the general solution takes the form

$$x_t = c_1 \lambda_1^t + \cdots + c_m \lambda_m^t,$$

where c_i, $i = 1, \ldots, m$ are arbitrary.

For the other cases, general solutions depend on the existence of repeated or complex conjugate eigenvalues. If there is a real eigenvalue λ_1 of multiplicity k, then k linearly independent solutions can be formed by multiplying by powers of t:

$$\lambda_1^t, t\lambda_1^t, t^2\lambda_1^t, \ldots, t^{k-1}\lambda_1^t.$$

If there are complex eigenvalues $\lambda_{1,2} = r(\cos\phi \pm i \sin\phi)$ of multiplicty k, then there are $2k$ linearly independent solutions:

$$r^t \cos(t\phi), r^t \sin(t\phi), tr^t \cos(t\phi), tr^t \sin(t\phi), \ldots, t^{k-1} r^t \cos(t\phi), t^{k-1} r^t \sin(t\phi)$$

5.3.2 Nonhomogeneous equations

An $m^{th}-$order linear nonhomogeneous equation is defined by

$$a_0 x_{t+m} + a_1 x_{t+m-1} + \cdots + a_m x_t = b(t). \tag{5.7}$$

Theorem 5.3.7. *The general solution of* (5.7) *is*

$$x_t = x_t^p + \sum_{i=1}^{m} a_i x_t^i$$

where x_t^p is a particular solution of the nonhomogeneous equation and $\{x_t^1, x_t^2, \ldots, x_t^k\}$ is a fundamental set of solutions of the m^{th}-order homogeneous equation (5.6)*.*

To find a particular solution of a nonhomogeneous equation, there exist several methods:

Method of undetermined coefficient: making a guess as to the form of the particular solution, and then substituting this function in the difference equation. This method works if the nonhomogeneous term $b(t)$ is a linear combination or product of terms having one of the forms

$$a^t, \quad \cos(ct), \quad \sin(ct), \quad t^k.$$

See Table 5.1.

Method of variation of constants Detailler

$b(t)$	x_t^p
a^t	$c_1 a^t$
t^k	$c_0 + c_1 t + c_2 t^2 + \cdots + c_k t^k$
$t^k a^t$	$c_0 a^t + c_1 t a^t + c_2 t^2 a^t + \cdots + c_k t^k a^t$
$\sin(ct), \cos(ct)$	$c_1 \sin(ct) + c_2 \cos(ct)$
$a^t \sin(ct), a^t \cos(ct)$	$(c_1 \sin(ct) + c_2 \cos(ct))a^t$
$a^t t^k \sin(ct), a^t t^k \cos(ct)$	$(d_0 + d_1 t + d_2 t^2 + \cdots + d_k t^k) \sin(ct) a^t +$ $(c_0 + c_1 t + c_2 t^2 + \cdots + c_k t^k) \cos(ct) a^t$

Table 5.1: Particular solutions

5.3.3 Qualitative analysis

What is the long-term behaviour of the solutions?

For linear difference equations, the asymptotic behavior depends on the eigenvalues: real and complex and the magnitude of eigenvalues.

Definition 5.3.8. *Magnitude of an eigenvalue:*

- *If $\lambda = A$ is real, $|\lambda| = |A|$ is the absolute value.*
- *If $\lambda = A + iB$ is complex, $|\lambda| = |A + iB| = \sqrt{A^2 + B^2}$ is the modulus.*

Definition 5.3.9. *An eigenvalue λ_i such that*

$$|\lambda_i| \geq |\lambda_j|$$

for all $j \neq i$ is called the dominant *eigenvalue. If the inequality is strict, then λ_i is a* strictly dominant *eigenvalue.*

Let the general solution of (5.6) be

$$x_t = \sum_{i=1}^{m} c_i \lambda_i^t.$$

The limiting behavior of the general solution is determined by the behavior of the dominant solution (corresponding to the dominant eigenvalue). Suppose that there exists a strictly dominant eigenvalue λ_1, then

$$x_t = \lambda_1^t \left[c_1 + \sum_{i=2}^{m} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^t \right].$$

Since $\left| \frac{\lambda_i}{\lambda_1} \right| < 1$, for all $i \neq 1$, it follows that $\left\{ \left(\frac{\lambda_i}{\lambda_1} \right)^t \right\}$ is a geometric sequence with general term $\left| \frac{\lambda_i}{\lambda_1} \right| < 1$, and thus

$$\left(\frac{\lambda_i}{\lambda_1} \right)^t \to 0 \quad \text{as } t \to +\infty.$$

Then

$$\lim_{t \to +\infty} x_t = \lim_{t \to +\infty} c_1 \lambda_1^t.$$

Depending on the value of λ_1 there are different situations. First, suppose $\lambda_1 \in \mathbb{R}$. Then

- if $\lambda_1 > 1$, then $\lim_{t\to+\infty} c_1\lambda_1^t = \infty$ (monotonically diverges $\Rightarrow$ unstable system),
- if $\lambda_1 = 1$, then the solution is constant,
- if $0 < \lambda_1 < 1$: $\lim_{t\to+\infty} c_1\lambda_1^t = 0$ (monotonically decreases to 0 $\Rightarrow$ stable system),
- if $-1 < \lambda_1 < 0$: $\lim_{t\to+\infty} c_1\lambda_1^t = 0$ (oscillating around zero and converging to 0 $\Rightarrow$ stable system),
- if $\lambda_1 = -1$, the system oscillates between two values c_1 and $-c_1$,
- and if $\lambda_1 < -1$, the system is oscillating but increasing in magnitude (unstable system).

Suppose now that $\lambda_1 \in \mathbb{C}$. Then

- if $|\lambda_1| > 1$, the system oscillates but increases in magnitude (unstable system),
- if $|\lambda_1| = 1$, the system oscillates but constant magnitude,
- if $|\lambda_1| < 1$, the system oscillates but converges to 0 (stable system).

The magnitude of the eigenvalues determines whether solutions are unbounded or bounded. The nature (real or complex) determines whether solutions oscillate or not.

In the case of a nonhomogeneous difference equation with a constant nonhomogeneous term, if the system converges, it will converge to the equilibrium point x^* (not to 0 as previously).

5.4 First-order linear systems

A higher-order linear difference equation can be converted to a first-order linear system. Consider the m^{th}-order linear nonhomogeneous equation

$$a_0x_{t+m} + a_1x_{t+m-1} + \cdots + a_mx_t = b(t).$$

5.4.1 Generality of first-order systems

For convenience, x_t is now denoted $x(t)$. Let $Y(t)$ be an m-vector, $Y(t) = (y_1(t), y_2(t), \ldots, y_m(t))$, which satisfies

$$\begin{aligned} y_1(t) &= x(t) \\ y_2(t) &= x(t+1) \\ y_3(t) &= x(t+2) \\ &\vdots \\ y_m(t) &= x(t+m-1). \end{aligned}$$

The first element $y_1(t)$ is the solution $x(t)$. Hence a first-order difference equation in y is

$$\begin{aligned} y_1(t+1) &= y_2(t) \\ y_2(t+1) &= y_3(t) \\ y_3(t+1) &= y_4(t) \\ &\vdots \\ y_{m-1}(t+1) &= y_m(t) \\ y_m(t+1) &= -a_1 y_m(t) - \cdots - a_{m-1} y_2(t) - a_m y_1(t) + b(t) \end{aligned}$$

In matrix form,

$$Y(t+1) = AY(t) + B$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_m & -a_{m-1} & -a_{m-2} & \dots & -a_1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix}.$$

The matrix A has 1's along the superdiagonal and has the coefficients of the higher-order difference equation $-a_i$ (but with the signs reversed) along the last row. The matrix A thus defined is called the **companion matrix** of the $m^{th}-$order difference equation.

5.4.2 Solutions of linear systems

A solution to a first-order linear difference system $X(t+1) = AX(t) + B$ is the superposition of two solutions: the general solution X_h to the homogeneous system $X_h(t+1) = AX_h(t)$ and a particular solution X_p to the nonhomogeneous system $X_p(t+1) = AX_p(t) + B$. The general solution to the nonhomogeneous system is

$$X(t) = X_h(t) + X_p(t).$$

The homogeneous system has m linearly independent solutions: there are some direct and indirect methods to find these linearly independent solutions.

Indirect methods use the fact that the solution can be expressed as $X(t) = A^t X(0)$. In [4], methods to compute A^t are presented, then the general solution can be known.

Direct method to solve $X(t+1) = AX(t)$ where $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ is an $m \times m$ constant matrix: Assume that a solution has the following form $X(t) = \lambda^t V$ where V is an nonzero $m-$column vector and λ is a constant. Substituting $\lambda^t V$ into the linear system gives

$$\lambda^{t+1} V = A\lambda^t V,$$

then

$$(A - \lambda I)V = \mathbf{0} \tag{5.8}$$

where I is the $m \times m$ identity matrix and $\mathbf{0}$ is the zero vector. The zero solution $V = 0$ is the trivial solution; and (5.8) has an unique solution if $\det(A - \lambda I) \neq 0$. Hence, nonzero solutions V are obtained if and only if $(A - \lambda I)$ is singular if and only if

$$\det(A - \lambda I) = 0. \tag{5.9}$$

Equation (5.9) is referred as the **characteristic equation of matrix** A. The m solutions λ_i, $i = 1, \ldots, m$ of (5.9) are called the **eigenvalues** of the matrix A. The nonzero solutions V_i are the **eigenvectors** corresponding to the eigenvalue λ_i that are found by solving $(A-\lambda_i I)V_i = \mathbf{0}$.

Then the general solution is a linear combination of m linearly independent solutions $X_i = \lambda_i^t V_i$, $i = 1, \ldots, m$:

$$X(t) = \sum_{i=1}^{m} c_i \lambda_i^t V_i \tag{5.10}$$

where c_i are arbitrary constants.

Understanding the asymptotic behavior of the solution (5.10) does not require the knowlegde of the eigenvectors. The asymptotic behavior is determined by the eigenvalues and their magnitude. As the solution of $X(t+1) = AX(t)$ is $X(t) = A^t X(0)$, it follows from Theorem B.2.3 that $\lim_{t\to+\infty} X(t) = 0$ when $\rho(A) < 1$.

5.5 Fixed points

For that, as in the continuous time case, we first seek equilibrium solutions, that is, solutions for which no variation occurs. Because of the type of equation that arises when seeking such solutions, equilibrium solutions are usually called fixed points, in the context of discrete-time systems.

Definition 5.5.1 (Fixed point)**.** *Let f be a function. A point p such that $f(p) = p$ is called a* ***fixed point*** *of f.*

The existence of fixed points is guaranteed in a relatively general situation by the following two theorems.

Theorem 5.5.2. *Consider the closed interval $I = [a, b]$. If $f : I \to I$ is continuous, then f has a fixed point in I.*

Theorem 5.5.3. *Let I be a closed interval and $f : I \to \mathbb{R}$ be a continuous function. If $f(I) \supset I$, then f has a fixed point in I.*

Definition 5.5.4 (Periodic point)**.** *Let f be a function. If there exists a point p and an integer n such that*

$$f^n(p) = p, \quad \textit{but} \quad f^k(p) \neq p \textit{ for } k < n,$$

then p is a periodic point of f with (least) period n (or a n-periodic point of f).

Thus, p is a n-periodic point of f iff p is a 1-periodic point of f^n.

5.5.1 Local stability of fixed points and periodic points

Theorem 5.5.5. *Let f be a continuously differentiable function (that is, differentiable with continuous derivative, or C^1), and p be a fixed point of f.*

i) If $|f'(p)| < 1$, then there is an open interval $\mathcal{I} \ni p$ such that $\lim_{k\to\infty} f^k(x) = p$ for all $x \in \mathcal{I}$.

ii) If $|f'(p)| > 1$, then there is an open interval $\mathcal{I} \ni p$ such that if $x \in \mathcal{I}$, $x \neq p$, then there exists k such that $f^k(x) \notin \mathcal{I}$.

Definition 5.5.6. *Suppose that p is a n-periodic point of f, with $f \in C^1$.*

- *If $|(f^n)'(p)| < 1$, then p is an **attracting** periodic point of f.*
- *If $|(f^n)'(p)| > 1$, then p is an **repelling** periodic point of f.*

5.5.2 Bifurcations

Consider an equation

$$x_{t+1} = f_r(x_t), \tag{5.11}$$

where r is a parameter in $\mathbb{R}$. The function f_r is called a **parametrized family** of functions.

Definition 5.5.7 (Bifurcation)**.** *Let f_μ be a parametrized family of functions. Then there is a **bifurcation** at $\mu = \mu_0$ (or μ_0 is a bifurcation point) if there exists $\varepsilon > 0$ such that, if $\mu_0 - \varepsilon < a < \mu_0$ and $\mu_0 < b < \mu_0 + \varepsilon$, then the dynamics of $f_a(x)$ are "different" from the dynamics of $f_b(x)$.*

An example of "different" would be that f_a has a fixed point (that is, a 1-periodic point) and f_b has a 2-periodic point.

5.5.3 Global stability

The results presented here come from [1]. We use the words *equilbrium* and *fixed point* interchangeably. Throughout this Section, we consider the discrete time scalar equation

$$x_{t+1} = f(x_t) \tag{5.1}$$

induced by the mapping f.

Definition 5.5.8 (Globally attractive fixed point)**.** *Suppose that $\bar{x}$ is a fixed point of f, where $f : [0, a) \to [0, a)$, $0 < a \leq \infty$. Then $\bar{x}$ is said to be* globally attractive *if for all initial conditions $x_0 \in (0, a)$,*

$$\lim_{t \to \infty} x_t = \bar{x}.$$

Definition 5.5.9 (Globally asymptotically stable fixed point)**.** *The fixed point is said to be* globally asymptotically stable *if $\bar{x}$ is globally attractive and if $\bar{x}$ is locally stable.*

Globally attractive equilibria are locally attractive, therefore globally asymptotically stable equilibria are locally asymptotically stable.

5.6 Nonlinear difference equations

Here, we study autonomous difference equations. The methods presented here also apply to the linear systems studied in the previous sections, although for linear systems, it is more efficient to use the specific tools that can be derived.

Recall that a difference equation of first order takes the form

$$x_{t+1} = f(x_t), \tag{5.1}$$

with $f : \mathbb{R}^n \to \mathbb{R}^n$ a function and $x_t \in \mathbb{R}^n$. Specific hypotheses will be made about f where needed. Starting from an initial point x_0, we have

$$\begin{aligned} x_1 &= f(x_0) \\ x_2 &= f(x_1) = f(f(x_0)) = f^2(x_0) \\ x_3 &= f(x_2) = f(f(f(x_0))) = f^3(x_0) \\ &\dots \\ x_t &= f(f(f(\dots f(x_0)))) = f^t(x_0) \end{aligned}$$

where the superscript t is the number of time steps or iterations beginning from the initial value x_0. $f^m(x_0)$ is called the m^{th} **iterate** of f.

In this section, we will repeatedly go from $f : \mathbb{R}^n \to \mathbb{R}^n$ to $f : \mathbb{R} \to \mathbb{R}$. This will be made explicit in the statement of results. In general, if no domain and range are specified for f, it is assumed that they are $\mathbb{R}^n$.

5.6.1 Equilibrium solution - Periodic solution

Definition 5.6.1. *A point x^* in the domain of f is said to be an **equilibrium point** (an equilibrium solution) of the first-order difference equation* (5.1) *if it is a fixed point of f* i.e. *a constant solution that satisfies*

$$f(x^*) = x^*.$$

Graphically, if $f : \mathbb{R} \to \mathbb{R}$, an equilibrium point is the $x-$coordinate of a point where the graph of $f(x)$ intersects the diagonal $y = x$ (since at such a point, $x = f(x)$). If $f : \mathbb{R}^2 \to \mathbb{R}^2$, we often write the system as

$$\begin{aligned} x_{t+1} &= f(x_t, y_t) \\ y_{t+1} &= g(x_t, y_t), \end{aligned}$$

and an equilibrium solution is then a point $(\bar{x}, \bar{y})$ such that $\bar{x} = f(\bar{x}, \bar{y})$ and $\bar{y} = g(\bar{x}, \bar{y})$.

Remark – Equilibrium solutions are interesting because they represent the "resting states", the "stationary states" of the system. No change occurs from step t to step $t + 1$. ∘

Definition 5.6.2. *A **periodic solution** of (least) period $m > 1$ of the difference equation* (5.1) *is a real-valued solution $\bar{x}_k$ satisfying*

$$f^m(\bar{x}_k) = \bar{x}_k$$

and

$$f^i(\bar{x}_k) \neq \bar{x}_k \quad i = 1, 2, \dots, m-1.$$

Note that an equilibrium point is a solution of period 1. Graphically, for $f : \mathbb{R} \to \mathbb{R}$, a periodic point is the $x-$coordinate of a point at which the graph of $f^m(x)$ intersects the diagonal line $y = x$.

Definition 5.6.3. *An* **m-*cycle*** *is a set of points* $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m\}$ *where for each* $k = 1, \ldots, m$, $\bar{x}_k$ *is a periodic solution of period* m. *The set* $\{\bar{x}_1, f(\bar{x}_1), \ldots, f^{m-1}(\bar{x}_1)\}$ *is called a* ***periodic orbit*** *of* $\bar{x}_1$

5.6.2 Local stability in first-order equations

Local stability of an equilibrium solution implies that solutions approach the equilibrium only if they are initially close to it. Global stability of an equilibrium is much stronger: global stability implies that regardless of the initial condition, solutions approach the equilibrium.

An equilibrium is called locally asymptotically stable if for any small perturbation away from the equilibrium, the solution returns to the equilibrium value.

Definition 5.6.4 (Local stability)**.** *A equilibrium solution* $\bar{x}$ *of* (5.1) *is* ***locally stable*** *if, for any* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that if* $\|x_0 - \bar{x}\| < \delta$, *then*

$$\|x_t - \bar{x}\| = \|f^t(x_0) - \bar{x}\| < \varepsilon, \quad \forall t > 0.$$

If a fixed point $\bar{x}$ *is not stable, then it is* ***unstable****.*

Definition 5.6.5 (Local attractivity)**.** *A equilibrium solution* $\bar{x}$ *is* ***locally attracting*** *if there exists* $\eta > 0$ *such that*

$$\|x_0 - \bar{x}\| < \eta \quad \textit{implies} \quad \lim_{t\to\infty} x_t = \bar{x}.$$

If $\eta = \infty$, *then* $\bar{x}$ *is a* global attractor *(or is* globally attracting*).*

Definition 5.6.6 (Local asymptotic stability)**.** *The equilibrium solution* $\bar{x}$ *is* ***locally asymptotically stable*** *if it is locally stable and locally attracting.*

The convergence behavior for a first-order difference equation that is locally asymptotically stable may take the form of convergent oscillations or convergent exponential solutions. If the solution values tend to amplify and do not converge to the equilibrium, the equilibrium is unstable. Such instability may appear as divergent oscillations or divergent exponential solutions. When the equilibrium is stable but not asymptotically stable it is said *neutrally stable*.

Consider an equilibrium solution $\bar{x}$ of (5.1). Define a new variable

$$u_t = x_t - \bar{x},$$

which represents a small quantity called a **perturbation** of the equilibrium solution. Then u_t satisfies the difference equation (5.1)

$$u_{t+1} = x_{t+1} - \bar{x} = f(x_t) - \bar{x} = f(u_t + \bar{x}) - f(\bar{x}) = g(u_t)$$

where $g(u) = f(u + \bar{x}) - f(\bar{x})$.

Note that zero is a fixed point of g if and only if $\bar{x}$ is a fixed point of f. In addition, zero is a locally stable (unstable, or locally asymptically stable) fixed point of g if and only if $\bar{x}$ is a locally stable (unstable or locally asymptotically stable) fixed point of f. To determine the stability of $\bar{x}$, we assume that f has a second order derivative in some interval I containing $\bar{x}$. Then, by Taylor's approximation,

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \frac{f''(\varepsilon)}{2!}(x - \bar{x})^2 + \text{h.o.t.},$$

for $\varepsilon \in I$, where h.o.t. denotes higher order terms. Thus, neglecting terms of order 2 or higher, for $(x - \bar{x})$ sufficiently small, we have the linear approximation

$$\underbrace{f(x_t) - \bar{x}}_{u_{t+1}} = f'(\bar{x}) \underbrace{(x_t - \bar{x})}_{u_t}.$$

So

$$u_{t+1} = f'(\bar{x})u_t \tag{5.12}$$

is referred to as the *linear approximation* to the difference equation (5.1) at the equilibrium $\bar{x}$.

If the initial condition is sufficiently close to $\bar{x}$, then the dynamics of u_t is determined by the linearization (5.12). In term of perturbation, to understand whether small perturbations u_t from the equilibrium solution increase or decrease, we can solve (5.12) by using the method for linear difference equations. We know that the solution of (5.12) will be decreasing whenever $|f'(\bar{x})| < 1$. Therefore, the value of $f'(\bar{x})$ determines whether $\bar{x}$ is locally asymptotically stable or unstable.

Theorem 5.6.7 (Condition for stability)**.** *Assume f' is continuous on an open interval I containing $\bar{x}$ and $\bar{x}$ is a fixed point of f. Then $\bar{x}$ is a locally asymptotically stable equilibrium of* (5.1) *if*

$$|f'(\bar{x})| < 1$$

and unstable if

$$|f'(\bar{x})| > 1$$

Definition 5.6.8. *An equilibrium $\bar{x}$ of* (5.1) *is said to be* ***hyperbolic*** *if $|f'(\bar{x})| \neq 1$. Otherwise ($|f'(\bar{x})| = 1$), it is said to be* ***nonhyperbolic*** *.*

Theorem 5.6.9. *Suppose that $f'(\bar{x}) = 1$ for an equilibrium solution $\bar{x}$ of $x_{t+1} = f(x_t)$, and f''' is continuous on an open interval containing $\bar{x}$, then the following statement hold:*

- *$f''(\bar{x}) \neq 0$, then the $\bar{x}$ is unstable.*
- *$f''(\bar{x}) = 0$ and $f'''(\bar{x}) > 0$, then $\bar{x}$ is unstable.*
- *$f''(\bar{x}) = 0$ and $f'''(\bar{x}) < 0$, then $\bar{x}$ is locally asymptotically stable.*

Definition 5.6.10 (Schwarzian derivative)**.** *The* ***Schwarzian derivative*** *of a function f at x is denoted $(Sf)(x)$ and defined by*

$$(Sf)(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2.$$

Note that $f'(x) = -1$, $(Sf)(x) = -f'''(x) - \frac{3}{2}f''(x)^2$.

Theorem 5.6.11. *Suppose that $f'(\bar{x}) = -1$ for an equilibrium solution $\bar{x}$ of (5.1), and f''' is continuous on an open interval containing $\bar{x}$, then the following statement hold:*

- *$(Sf)(\bar{x}) > 0$, then the $\bar{x}$ is unstable.*
- *$(Sf)(\bar{x}) < 0$, then $\bar{x}$ is locally asymptotically stable.*

Theorem 5.6.12. *Suppose f' is continuous on an open interval I and the $m-$cycle*

$$\{\bar{x}_1, f(\bar{x}_1), \ldots, f^{m-1}(\bar{x}_1)\}$$

of the difference equation

$$x_{t+1} = f(x_t)$$

is contained in I. Then the $m-$cycle is locally asympotically stable if

$$\left|\frac{d[f^m(\bar{x}_k)]}{dx}\right| < 1$$

for some k and unstable if

$$\left|\frac{d[f^m(\bar{x}_k)]}{dx}\right| > 1$$

for some k.

Corollary 5.6.13. *Suppose $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m\}$ is an $m-$cycle of $x_{t+1} = f(x_t)$. Then the $m-$cycle is locally asymptotically stable if*

$$|f'(\bar{x}_1)f'(\bar{x}_2)\ldots f'(\bar{x}_m)| < 1$$

Example – $\{\bar{x}_1, \bar{x}_2\}$ is a $2-$cycle that is locally asymptotically stable if and only if

$$\left|\left.\frac{d[f]}{dx}\right|_{\bar{x}_1} \left.\frac{d[f]}{dx}\right|_{\bar{x}_2}\right| < 1.$$

◇

Cobwebbing method for first-order equation

Graphical method to answer qualitative questions about the solution of

$$x_{t+1} = f(x_t).$$

In the $(x_t x_{t+1})-$plane, sketch $x_{t+1} = x_t$ and $x_{t+1} = f(x_t)$:

- any intersections of these graphs is an equilibrium solution of the difference equation.
- to investigate the behavior of the solutions
  - choose a starting value x_0, and begin at the point (x_0, x_0) in the $(x_t x_{t+1})-$plane.
  - draw a vertical line to the curve $x_{t+1} = f(x_t)$; this reaches the curve at $(x_0, f(x_0)) = (x_0, x_1)$.
  - draw a horizontal line to the diagonal $x_{t+1} = x_t$; this reaches the diagonal at the point (x_1, x_1).
  - Repeat the process to arrive at (x_2, x_2) and indefinitely until the behavior of the equation with this starting value becomes clear.
  - If necessary, re-do the same with other starting values.

5.6.3 Global stability in first-order equations

Global stablity of an equilibrium removes the restrictions on the initial conditions. In global asympotic stability, solutions approach the equilibrium solution for all initial conditions.

Definition 5.6.14. *Suppose that $\bar{x}$ is an equilibrium solution of the difference equation*

$$x_{t+1} = f(x_t),$$

where $f : [0, a) \to [0, a)$, $0 < a \leq \infty$. Then $\bar{x}$ is said to be globally attractive if for all initial conditions $x_0 \in (0, a)$,

$$\lim_{t\to\infty} x_t = \bar{x}.$$

The equilibrium is said to be globally asymptotically stable if $\bar{x}$ is globally attractive and if $\bar{x}$ is locally stable.

Globally attractive equilibria are locally attractive, therefore globally asymptotically stable equilibria are locally asymptotically stable.

If f is a continuous map, global attractivity is equivalent to global asymptotic stability.

Theorem 5.6.15. *Suppose that for system* (5.1)*, the function f satisfies*

i) f is continuous on $[0, a)$, $0 < a \leq \infty$,

ii) $f : [0, a) \to [0, a)$, $0 < a \leq \infty$,

iii) $0 < f(x) < x$ for all $x \in [0, a)$.

Then the origin is globally asymptotically stable for (5.1)*.*

The following provide necessary and sufficient conditions for global asymptotic stability of a positive equilibrium $\bar{x}$.

Theorem 5.6.16. *The difference equation* (5.1)*, with f satisfying*

i) f is continuous on $[0, a)$, $0 < a \leq \infty$,

ii) $f : [0, a) \to [0, a)$, $0 < a \leq \infty$,

iii) $f(0) = 0$, $f(\bar{x}) = \bar{x}$,

iv) $f(x) > x$ for $0 < x < \bar{x}$,

v) $f(x) < x$ for $\bar{x} < x < a$,

vi) if f has a maximum at x_M in $(0, \bar{x})$, then f is decreasing for $x > x_M$,

has a globally asymptotically stable equilibrium at $\bar{x}$ if and only if f has no 2-cycles.

To prove that there are no 2-cycles, the next result is helpful.

Theorem 5.6.17. *Let f' be continuous on an interval I and $f : I \to I$. If $1 + f'(x) \neq 0$ for all $x \in I$ then* (5.1) *has no $2-$cycles in I.*

Theorem 5.6.18. *Consider the difference equation* (5.1)*. If f satisfies*

i) f *is continuous on* $[0, a)$, $0 < a \leq \infty$,

ii) $f : [0, a) \to [0, a)$, $0 < a \leq \infty$,

iii) $\bar{x} \in (0, a)$ *such that* $x < f(x) < \bar{x}$ *for* $0 < x < \bar{x}$ *and* $\bar{x} < f(x) < x$ *for* $x > \bar{x}$

then the difference equation (5.1) *has a globally asymptotically stable equilibrium at* $\bar{x}$.

Theorem 5.6.19. *Consider the difference equation* (5.1).

a. *Suppose that* f *satisfies*

i) f *is continuous on* $[0, a)$, $0 < a \leq \infty$,
ii) $f : [0, a) \to [0, a)$, $0 < a \leq \infty$,
iii) $f(0) = 0$, $f(\bar{x}) = \bar{x}$
iv) $f(x) > x$ *for* $0 < x < \bar{x}$
v) $f(x) < x$ *for* $\bar{x} < x < a$

but has no maximum in $(0, \bar{x})$. *Then* $\bar{x}$ *is globally asymptotically stable.*

b. *Suppose that* f *satisfies*

i) f *is continuous on* $[0, a)$, $0 < a \leq \infty$,
ii) $f : [0, a) \to [0, a)$, $0 < a \leq \infty$,
iii) $f(0) = 0$, $f(\bar{x}) = \bar{x}$
iv) $f(x) > x$ *for* $0 < x < \bar{x}$
v) $f(x) < x$ *for* $\bar{x} < x < a$
vi) if f *has a maximum at* x_M *in* $(0, \bar{x})$, *then* f *is decreasing for* $x > x_M$

has a maximun x_M *in* $(0, \bar{x})$. *Then* $\bar{x}$ *is globally asymptotically stable if and only if* $f^2(x) > x$ *for all* $x \in [x_M, \bar{x})$

5.6.4 Bifurcation diagrams

To summarize the range of behaviors, a diagram of bifurcation can be used by depicting the locations and the stability properties of periodic solutions. To illustrate the effect of a parameter variation on existence and stability properties of periodic solutions.

Dependence of a difference equation on a parameter can be noted

$$x_{t+1} = f(x_t, r).$$

The values of r where the behavior changes are known as the *bifurcation values* and the points $(r, \bar{x}(r))$ are the *bifurcation points* with $\bar{x}(r)$ is the value of periodic solutions for the parameter value r.

A change in the solution behavior occurs when an equilibrium or a $m-$cycle changes stability.

- on the horizontal axis, the parameter value r.

- on the vertical axis, the magnitudes of equilibrium solutions or cycles.
- an unstable equilibrium or cycle is denoted by a dashed curve.
- a stable equilibrium or cycle is denoted by a solid curve.

Definition 5.6.20. *Deterministic chaos is a pattern of fluctuations that may seem to be stochastic but it is actually produced in a deterministic manner, by autonomous nonlinear dynamic processes.*

A property of the chaotic system is the extrem sensitivity to initial conditions.

Sensitive dependence to initial conditions means that a small perturbation in these initial conditions will grow exponentially with time. Even if, theoretically, it should be possible to predict the future dynamic as a function of time, in reality it is impossible, because the smallest error in the specification of the initial state leads to a great error in future predictions. In fact, in deterministic chaos, the knowledge of the state of the system during as long a time as we want, doesn't allow us to predict its further evolution [6].

A chaotic system has cycles of every period.

5.7 Systems of nonlinear equations

Consider the system

$$\begin{aligned} x_{t+1} &= f(x_t, y_t) \\ y_{t+1} &= g(x_t, y_t), \end{aligned} \tag{5.13}$$

where f and g are nonlinear function. An equilibrium $(\bar{x}, \bar{y})$ to (5.13) satisfies the fixed point problem

$$\begin{aligned} \bar{x} &= f(\bar{x}, \bar{y}) \\ \bar{y} &= g(\bar{x}, \bar{y}). \end{aligned}$$

Assume that a point $(\bar{x}, \bar{y})$ has been found that satisfies this fixed point problem. What is the stability of the equilibrium $(\bar{x}, \bar{y})$?

The first step consists in the linearization of the system about the equilibrium. To linearize the system, we use the Taylor series expansion of functions of two variables to approximate f and g about $(\bar{x}, \bar{y})$. For f, the Taylor series is given by

$$f(x,y) = f(\bar{x},\bar{y}) + \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}} (x-\bar{x}) + \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}} (y-\bar{y}) + \left.\frac{\partial^2 f}{\partial x^2}\right|_{\bar{x},\bar{y}} \frac{(x-\bar{x})^2}{2!} + \left.\frac{\partial^2 f}{\partial y^2}\right|_{\bar{x},\bar{y}} \frac{(y-\bar{y})^2}{2!} + \dots$$

We consider only terms of degree 0 and 1, discarding all quadratic and higher order terms. Note that this defines the equation of the plane tangent to the surface $f(x, y)$ at the point $(\bar{x}, \bar{y})$. Thus, for f,

$$f(x,y) = f(\bar{x},\bar{y}) + \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}} (x-\bar{x}) + \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}} (y-\bar{y}).$$

The Taylor series for g is obtained similarly.

Now consider a small perturbation $u = x - \bar{x}$ and $v = y - \bar{y}$ about the fixed point $(\bar{x}, \bar{y})$. Then

$$f(x,y) = f(\bar{x},\bar{y}) + \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}} u + \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}} v$$

and

$$g(x,y) = g(\bar{x},\bar{y}) + \left.\frac{\partial g}{\partial x}\right|_{\bar{x},\bar{y}} u + \left.\frac{\partial g}{\partial y}\right|_{\bar{x},\bar{y}} v.$$

Therefore,

$$f(x,y) - \bar{x} = \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}} u + \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}} v$$

$$g(x,y) - \bar{y} = \left.\frac{\partial g}{\partial x}\right|_{\bar{x},\bar{y}} u + \left.\frac{\partial g}{\partial y}\right|_{\bar{x},\bar{y}} v.$$

As $u_t = x_t - \bar{x}$, $v_t = y_t - \bar{y}$ and $u_{t+1} = x_{t+1} - \bar{x}$, $v_{t+1} = y_{t+1} - \bar{y}$, it follows that $u_{t+1} = f(x_t, y_t) - \bar{x}$ and $v_{t+1} = g(x_t, y_t) - \bar{y}$. Therefore

$$u_{t+1} = f(x_t,y_t) - \bar{x} = \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}} u_t + \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}} v_t$$

$$v_{t+1} = g(x_t,y_t) - \bar{y} = \left.\frac{\partial g}{\partial x}\right|_{\bar{x},\bar{y}} u_t + \left.\frac{\partial g}{\partial y}\right|_{\bar{x},\bar{y}} v_t.$$

Writing this in vector form, we have that the linearization of (5.13) about the equilibrium $(\bar{x}, \bar{y})$ where $u_t = x_t - \bar{x}$ and $v_t = y_t - \bar{y}$ is

$$V_{t+1} = JV_t,$$

where $V_t = (u_t, v_t)^T$ and J is the Jacobian of $(f, g)^T$ evaluated at $(\bar{x}, \bar{y})$,

$$J = \begin{pmatrix} \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}} & \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}} \\ \left.\frac{\partial g}{\partial x}\right|_{\bar{x},\bar{y}} & \left.\frac{\partial g}{\partial y}\right|_{\bar{x},\bar{y}} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The eigenvalues of the Jacobian J determine the local stability of the nonlinear system: if they satify $|\lambda_i| < 1$ for $i = 1, 2$, i.e, if the spectral radius $\rho(J) < 1$, then from Theorem **??**, it follows that $\lim_{t\to\infty} J^t = 0$. In that case, we have $\lim_{t\to\infty} V_t = 0$. This, in turn, implies that

$$\lim_{t\to\infty} f(x_t, y_t) = \bar{x}$$
$$\lim_{t\to\infty} g(x_t, y_t) = \bar{y},$$

that is,

$$\lim_{t\to\infty} x_t = \bar{x}$$
$$\lim_{t\to\infty} y_t = \bar{y}.$$

Since Theorem **??** gives a necessary and sufficient condition for a matrix A to have iterates going to the zero matrix, it also follows that if one of the eigenvalues has modulus larger than or equal to 1, then V_t does not converge, implying that $(\bar{x}, \bar{y})$ is unstable.

To find the eigenvalues of J we need to solve

$$\det(J - \lambda I) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix} = 0.$$

The characteristic equation is

$$\lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a12a_{21} = \lambda^2 - \mathrm{tr}(J)\lambda + \det(J) = 0,$$

and thus the eigenvalues are

$$\lambda_{1,2} = \frac{\mathrm{tr}(J) \pm \sqrt{\mathrm{tr}(J)^2 - 4\det(J)}}{2}.$$

Theorem 5.7.1. *Let $f(x, y)$ and $g(x, y)$ be two functions with continuous first-order partial derivatives in x and y on some set containing $(\bar{x}, \bar{y})$. Then the equilibrium $(\bar{x}, \bar{y})$ of the nonlinear system*

$$\begin{aligned} x_{t+1} = &\quad f(x_t, y_t) \\ y_{t+1} = &\quad g(x_t, y_t) \end{aligned}$$

is locally asymptotically stable if the eigenvalues of the Jacobian matrix J evaluated at the equilibrium $(\bar{x}, \bar{y})$ satisfy $|\lambda_i| < 1$ if and only if

$$|tr(J)| < 1 + \det(J) < 2.$$

The equilibrium is unstable if some $|\lambda_i| > 1$, that is, if any one of three inequalities is satisfied

- $tr(J) > 1 + \det(J)$,
- *or* $tr(J) < -1 - \det(J)$,
- *or* $\det(J) > 1$

See Figure **??** for illustration

Higher-order difference equations

Local stability criteria for first or higher order difference equation depend on the behavior of the linearization of the system.

Consider a first-order system of n nonlinear equations $X(t) = (x_1(t), x_2(t), x_3(t), \ldots, x_n(t))^T$

$$X(t + 1) = F(X(t))$$

where $F = (f_1, f_2, f_3, \ldots, f_n)^T$ with $f_i = f_i(x_1, x_2, x_3, \ldots, x_n)$ for $i = 1, 2, 3, \ldots, n$. The point $\bar{X}$ is an equilibrium of the $n-$dimensional nonlinear system.

If $U(t) = X(t) - \bar{X}$, the linearization of the $n-$dimensional nonlinear system about $\bar{X}$ is

$$U(t) = JU(t)$$

where J is the Jacobian matrix evaluated at $\bar{X}$

$$J = \begin{pmatrix} \frac{\partial f_1(\bar{X})}{\partial x_1} & \frac{\partial f_1(\bar{X})}{\partial x_2} & \cdots & \frac{\partial f_1(\bar{X})}{\partial x_n} \\ \frac{\partial f_2(\bar{X})}{\partial x_1} & \frac{\partial f_2(\bar{X})}{\partial x_2} & \cdots & \frac{\partial f_2(\bar{X})}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_n(\bar{X})}{\partial x_1} & \frac{\partial f_n(\bar{X})}{\partial x_2} & \cdots & \frac{\partial f_n(\bar{X})}{\partial x_n} \end{pmatrix}$$

The eigenvalues of the Jacobian J determine the local stability of the $n-$dimensional nonlinear system. Eigenvalues are solutions of the characteristic equation

$$\det(J - \lambda I) = 0.$$

The eigenvalues λ are the zeros of the following n^{th}degree characteristic equation

$$p(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + a_3\lambda^{n-3} + \cdots + a_n \tag{5.14}$$

The conditions that must be satisfied for local asymptotic stability are known as the Jury Conditions or Schur-Cohn Criteria: they ensure that $|\lambda_i| < 1$.

Theorem 5.7.2 (Jury conditions or Schur-Cohn Criteria, for $n = 3$)**.** *Consider the characteristic polynomial*

$$p(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + a_3.$$

The solutions λ_i, $i = 1, 2, 3$, of $p(\lambda) = 0$ satisfy $|\lambda_i| < 1$ if and only if the following three conditions hold:

i) $p(1) = 1 + a_1 + a_2 + a_3 > 0$,

ii) $(-1)^3 p(-1) = 1 - a_1 + a_2 - a_3 > 0$

iii) $1 - (a_3)^2 > |a_2 - a_3 a_1|$

Some necessary conditions for $|\lambda_i| < 1$:

Theorem 5.7.3. *If the solutions λ_i, $i = 1, 2, \ldots, n$ of*

$$p(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + a_3\lambda^{n-3} + \cdots + a_n = 0$$

satisfy $|\lambda_i| < 1$ then

- $p(1) > 0$
- $(-1)^n p(-1) > 0$
- $|a_n| < 1$

Chapter 6

Deterministic discrete time models

In Section 3.5, the discrete time logistic equation was studied, and some theory for difference equations was presented in Chapter 5. Here, we give several examples, more or less detailed, of the use of discrete time systems.

The population dynamics of single species with seasonal reproduction and first-order feedback are often modelled using a single difference equation, so our first few models are of this type.

6.1 Other applications of the logistic map

Here, we list a few contexts in which the logistic map has been used, besides the growth of the human population.

6.1.1 Tumor cell growth

A population of tumor cells $N(t)$ growing in a container can be modeled using a logistic map. Here, r is the rate of growth of the tumor cells. Normalization of $N(t)$ means that $N(t)$ represents the fraction of the total population of cells contained in the cell culture. The cell culture can support a maximal number of cells represented by 1. The main assumption of the model is that the growth rate is constant.

6.2 Bacteria population

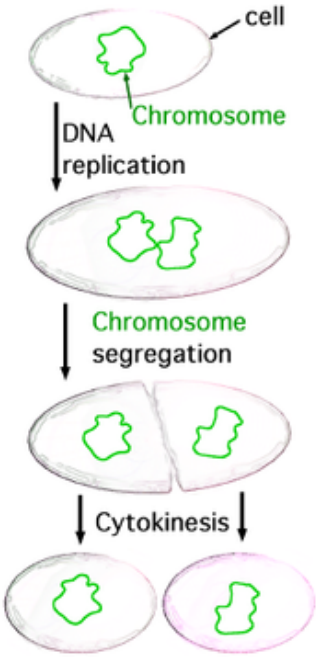
E. coli are able to divide every 20 minutes under optimal conditions. Describe the temporal evolution of a colony of E. coli.

Definition 6.2.1. *Cell division is the process by which a cell, called the parent cell, divides into two cells, called daughter cells. Cell division is usually a small segment of a larger cell cycle.*

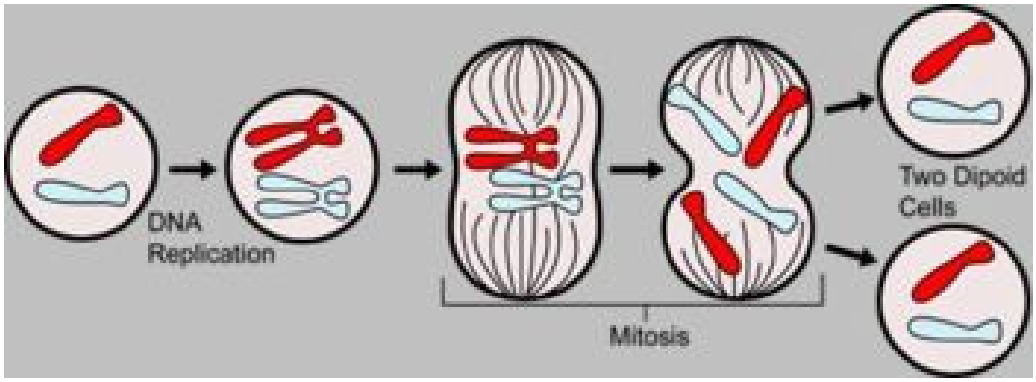
- *Prokaryotic cells: binary fission*
- *Eukaryotic cells: mitosis+cytokinesis*

Definition 6.2.2. *Binary fission: The prokaryotic chromosome is a single DNA molecule that first replicates, then attaches each copy to a different part of the cell membrane. When the cell begins to pull apart, the replicate and original chromosomes are separated. Following cell*

splitting (cytokinesis), there are then two cells of identical genetic composition (except for the rare chance of a spontaneous mutation).



Definition 6.2.3. *Mitosis+Cytokinesis: Mitosis is the process by which a cell separates its duplicated genome into two identical halves. It is generally followed immediately by cytokinesis which divides the cytoplasm and cell membrane. This results in two identical daughter cells with a roughly equal distribution of organelles and other cellular components. Mitosis and cytokinesis together is defined as the mitotic (M) phase of the cell cycle, the division of the mother cell into two daughter cells, each the genetic equivalent of the parent cell.*



Organisms that reproduce by binary fission (asexual reproduction) exhibit exponential growth. If organisms (or cells) are synchronized, the formalism of difference equation can be used. The model (for an unlimited environment) to describe the temporal evolution of the E. Coli colony is expressed as

$$x_{t+1} = 2x_t, \quad t = 0, 1, 2, \ldots$$

where x_t is the state variable that represents the number of cells at the generation t. If the initial population x_0 is known then the solution is unique and is

$$x_t = 2^t x_0, \quad t = 0, 1, 2, \ldots$$

The asymptotic behavior of the solution is

$$\lim_{t\to\infty} x_t = \infty$$

as $a = 2 > 1$. Note: If differential equation formalism was used

$$\frac{dx}{dt} = \frac{1440}{20}x$$

the solution would be $x(t) = x_0 e^{\frac{1440}{20}t}$.

6.3 The Ricker model

another model for describing a population $N(t)$ in a limited environment

$$N(t+1) = N(t)\exp\left\{r\left(1 - \frac{N(t)}{K}\right)\right\} = f(N(t)),$$

where r is the intrinsic growth rate and K is the carrying capacity. The growth rate $f(N(t))$ is increasing in $N(t)$ and the per capita growth $\frac{f(N)}{N}$ is decreasing in $N(t)$. The increase in population is not sufficient to compensate for the decrease in the per capita growth, then $\lim_{N(t)\to+\infty} f(N(t)) = 0$. Then the Ricker model can be referred as to overcompensatory.

- $r < 2$ Globally asymptotically stable equilibrium $\bar{x} = K$
- r = 2 Bifurcation into a stable 2-cycle
- r = 2.5 Bifurcation into a stable 4-cycle
- Then there is a series of cycle duplication: 8-cycle, 16-cycle, etc.
- r = 2.692 Chaos
- For $r > 2.7$ there are some regions where dynamics returns to a cycle, e.g., r=3.15.

6.4 The Hassell model

A population $N(t)$ in a limited environment

$$N(t+1) = \frac{rN(t)}{(1+N(t))^b}$$

where r is the intrinsic growth rate for small populations and b represents the inhibitive density-dependent feedback, usually attributed to the environment.

6.5 The Beverton-Holt model

$$N(t+1) = \frac{e^r K N(t)}{K + (e^r - 1)N(t)}$$

with r is the intrinsic growth rate, the carrying capacity is K.

6.6 Example of a 2-dimensional system

$$\begin{aligned} x(t+1) = &\quad x(t)(a - x(t) - y(t)), \quad a > 0 \\ y(t+1) = &\quad y(t)(b + x(t)), \quad 0 < b < 1. \end{aligned}$$

To find equilibria, solve the fixed point problem for x and y,

$$\begin{aligned} x = &\quad x(a - x - y) \\ y = &\quad y(b + x) \end{aligned}$$

We find 3 equilibria:

$$(\bar{x}_1, \bar{y}_1) = (0, 0), \quad (\bar{x}_2, \bar{y}_2) = (a - 1, 0), \quad (\bar{x}_3, \bar{y}_3) = (1 - b, a + b - 2).$$

The Jacobian of the system is

$$J = \begin{pmatrix} a - 2x - y & -x \\ y & b + x \end{pmatrix}.$$

The Jacobians evaluated at each equilibrium are:

$$J_{(\bar{x}_1,\bar{y}_1)} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad J_{(\bar{x}_2,\bar{y}_2)} = \begin{pmatrix} 2 - a & 1 - a \\ 0 & b + a - 1 \end{pmatrix} \quad J_{(\bar{x}_3,\bar{y}_3)} = \begin{pmatrix} b & -1 + b \\ a + b - 2 & 1 \end{pmatrix}$$

- Eigenvalues of $J_{(\bar{x}_1,\bar{y}_1)}$ are a and b. By definition $|b| < 1$. If $a < 1$, $(\bar{x}_1, \bar{y}_1)$ is L.A.S.
- $(\bar{x}_2, \bar{y}_2) = (a - 1, 0)$ exists only if $1 < a$. Eigenvalues of $J_{(\bar{x}_2,\bar{y}_2)}$ are $2 - a$ and $b + a - 1$, then the stability of $(\bar{x}_2, \bar{y}_2)$ depends on $|2 - a| < 1$ and $|b + a - 1| < 1$. These inequalities lead to $1 < a < 2 - b$.
- From Theorem 5.7.1, $(\bar{x}_3, \bar{y}_3)$ is L.A.S. if

  $$|1 + b| < 1 + b - (-1 + b)(a + b - 2) < 2.$$

  $(\bar{x}_3, \bar{y}_3)$ is positive if $b < 1$ and $a + b > 2$. Then the biological existence and the LAS of $(\bar{x}_3, \bar{y}_3)$ are possible for

  $$2 < a + b < 3.$$

6.7 An SIR epidemic model

Epidemic models were introduced in Section 4.4, and will be further discussed in Chapter 10 (in the ODE case). We refer in particular to Chapter 10, where much more details can be found.

Here, we present an example of SIR model in difference equation formalism.

Figure 6.1: SIR model: disease with recovery and permanent immunity, here birth=death=b.

$$S(t+1) = S(t) - \beta\frac{S(t)}{N}I(t) + b(I(t)+R(t)) \tag{6.1}$$

$$I(t+1) = (1-\gamma-b)I(t) + \beta\frac{S(t)}{N}I(t) \tag{6.2}$$

$$R(t+1) = R(t)(1-b) + \gamma I(t), \tag{6.3}$$

where $N = S(0) + I(0) + R(0)$. Parameters are β, the contact number (the average number of successful contacts made by one infected individual during the time t to $t+1$), $b = d$, the rate of birth and death, γ, the rate of recovery ($1/\gamma$ is the average length of the infectious period when there are no death, $1/(\gamma+b)$ is the average length of the infectious period when deaths are included).

It is easy to check that the total population is constant, $N = S(t) + I(t) + R(t)$. Also, solutions are nonnegative if $b, \gamma > 0$ and

$$0 < b + \gamma < 1, \qquad 0 < \beta < 1$$

.

We now consider the reduced system where $R(t) = N - I(t) - S(t)$ (which can be done since the total population is constant):

$$S(t+1) = S(t) - \beta\frac{S(t)}{N}I(t) + b(N - S(t))$$

$$I(t+1) = (1-\gamma-b)I(t) + \beta\frac{S(t)}{N}I(t).$$

We find two equilibria: the disease-free equilibrium $(S_1, I_1) = (N, 0)$ and the endemic equilibrium $(S_2, I_2) = \left(N\frac{(\gamma+b)}{\beta}, bN\frac{\beta-(\gamma+b)}{\beta(\gamma+b)}\right)$.

6.7.1 Stability of the disease free equilibrium (S_1, I_1)

The Jacobian evaluated at (S_1, I_1) is

$$J(S_1, I_1) = \begin{pmatrix} 1-b & -\beta \\ 0 & 1-b-\gamma+\beta \end{pmatrix}$$

as the Jacobian is upper triangular its eigenvalues are

$$\lambda_1 = 1-b, \quad \lambda_2 = 1-b-\gamma+\beta.$$

(S_1, I_1) is locally asymptotically stable if $|\lambda_{1,2}| < 1$

- from assumptions, we have $0 < \lambda_1 < 1$
- if $\frac{\beta}{\gamma+b} < 1$ (where $\mathcal{R}_0 = \frac{\beta}{\gamma+b}$ is the basis reproduction number), $0 < \lambda_2 < 1$

if $\mathcal{R}_0 < 1$, there exist only one (biologically plausible) equilibrium, the disease-free equilibrium, and it is L.A.S. (see Figure **??**)

6.7.2 Stability of the endemic equilibrium (S_2, I_2)

The Jacobian evaluated at (S_2, I_2) is

$$J(S_2, I_2) = \begin{pmatrix} 1 - b\mathcal{R}_0 & -\beta/\mathcal{R}_0 \\ b(\mathcal{R}_0 - 1) & 1 \end{pmatrix}$$

where $\mathrm{tr}(J(S_2, I_2)) = 2 - b\mathcal{R}_0$ (assume $\mathrm{tr}(J(S_2, I_2)) \geq 0$), and $\det(J(S_2, I_2)) = 1 - b\mathcal{R}_0 + \beta b(1 - \frac{1}{\mathcal{R}_0})$.

Condition for L.A.S (Theorem 5.7.1)

$$2 - b\mathcal{R}_0 < 2 - b\mathcal{R}_0 + \beta b(1 - \frac{1}{\mathcal{R}_0}) < 2$$

this condition is satified because

$$\beta(1 - 1/\mathcal{R}_0) < 1 < \mathcal{R}_0$$

If $1 < \mathcal{R}, \leq 2/b$, the endemic equilibrium exists and it is L.A.S. (see Figure **??**).

6.8 Predator-Prey models

Assumptions

- the prey has unlimited resources
- the prey's only threat is the predator
- the predator is a specialist; i.e., the predator's only food supply is the prey
- predator growth depends on the prey it catches

Variables

- $N(t)$ number of preys
- $P(t)$ number of predators

Parameters

- r intrinsic rate of growth of prey
- d rate of death of predators
- $eP(t)$ per capita prey reduction due to predation
- $bN(t)$ per capita predator increase due to prey

$$\begin{aligned} N(t+1) = &\ (1+r)N(t) - eN(t)P(t) \\ P(t+1) = &\ (1-d)P(t) + bN(t)P(t) \end{aligned}$$

Neubert and Kot model: If the prey follows a logistic growth

$$\begin{aligned} N(t+1) = &\ N(t) + rN(t)\left(1 - \tfrac{N(t)}{K}\right) - eN(t)P(t) \\ P(t+1) = &\ (1-d)P(t) + bN(t)P(t) \end{aligned}$$

6.9 Structured population models

Structured population models are used when the population can be organized or divided into various subclasses following traits such as age, life-stage or size. The variable that describes this trait is called the structuring variable.

- In age-structured model, the population is subdivided into age groups. For the human population, for example, age groups may be 5 year lengths, 0-5, 5-10, ..., or 1 year lengths.
- In stage-structured model, the population is organized into developmental stage: juveniles and adults, or for insects, egg, larva, pupa and adult.
- In size-structured model, individuals in the population are grouped according to size (length, weight, biomass).

The dynamic interactions among the stages, ages or sizes determine how the population structure changes over the time. Other structuring variables can be taken into account: sex and space.

6.10 Leslie matrix model

The Leslie Matrix (also called the Leslie Model) describes the growth of populations with structure (and their projected age distribution); the population is closed to migration and only one sex, usually the female, is considered.

Assume the population is closed to migration and only the females are modeled. Males are presented, but are not specifically modeled (when the sex ratio of males to females is a/b and the survival rate per age group is the same for males and females, then the number of males equals the number of females times a/b).

Let the total number of age groups is m (m the last reproductive age). During the interval of time t and $t+1$ individuals age from i to $i+1$: time interval coincides with the age interval.

- $x_i(t)$ number of females in the i^{th} age group at time t.
- b_i average number of newborn females produced by one females in the i^{th} age group that survive through the time interval in which they were born, $b_i \geq 0$
- s_i fraction of the i^{th} age group that lives to the $(i+1)st$ age, $0 < s_i \leq 1$

$$
\begin{aligned}
x_1(t+1) =& b_1x_1(t) + b_2x_2(t) + b_3x_3(t) + \dots b_mx_m(t) && \text{(6.4a)}\\
x_2(t+1) =& s_1x_1(t) && \text{(6.4b)}\\
x_3(t+1) =& s_2x_2(t) && \text{(6.4c)}\\
&\vdots && \text{(6.4d)}\\
x_m(t+1) =& s_{m-1}x_{m-1}(t) && \text{(6.4e)}\\
& && \text{(6.4f)}
\end{aligned}
$$

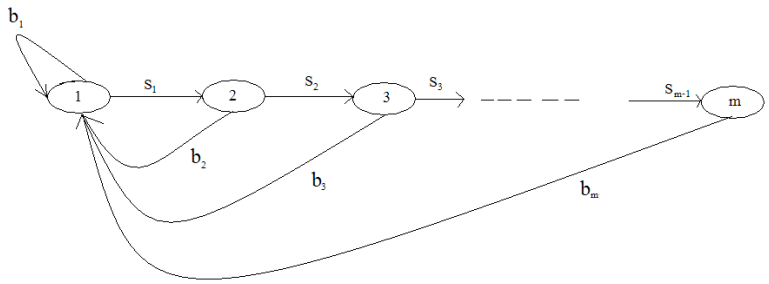


Figure 6.2: Life cycle graph of the Leslie matrix on m age classes: each node represents each age group x_i, and arcs represent relation between two groups. An arrow connects the node j to i if the ij^{th} element in the Leslie matrix L is nonzero.

Using matrix notation,

$$X(t+1) = \begin{pmatrix} x_1(t+1) \\ x_2(t+1) \\ x_3(t+1) \\ \vdots \\ x_m(t+1) \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & \dots & b_{m-1} & b_m \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & s_{m-1} & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_m(t) \end{pmatrix} = LX(t) \qquad (6.5)$$

where L is called the Leslie matrix: fertilities or fecundities on the first row and survival probabilities on the subdiagonal. All other entries in the Leslie matrix are zero.

$$X(1) = LX(0)$$
$$X(2) = LX(1) = L\left(LX(0)\right) = L^2X(0)$$

In general

$$X(t) = L^tX(0)$$

Definition 6.10.1. *A Leslie matrix is a nonnegative matrix.*

A necessary condition for a Leslie matrix to be irreductible is $b_m \neq 0$.

Frobenius Theorem gives sufficient conditions that guarantees the Leslie matrix has one positive strictly dominant eigenvalue.

If the Leslie matrix satisfies $L^p > 0$ for some positive integer p, then L is primitive. Then, the Frobenius Theorem states that in this case L has a unique strictly dominant eigenvalue λ_1 satisfying $|\lambda_1| > |\lambda_i|$, for $j \neq 1$, that is positive. Associated with the strictly dominant eigenvalue λ_1 is a positive eigenvector V_1, that is referred to as a stable age distribution.

Assume matrix L is irreductible and primitive and m eigenvectors form a linearly independent set; then the solution to

$$X(t+1) = LX(t)$$

can be written

$$X(t) = L^tX(0) = \sum_{i=1}^{m} c_i\lambda_i^tV_i$$

where λ_1 is the strictly dominant eigenvalue. Dividing the solution by λ_1^t gives

$$\frac{X(t)}{\lambda_1^t} = \frac{L^t X(0)}{\lambda_1^t} = c_1 V_1 + \frac{c_2 \lambda_2^t}{\lambda_1^t} V_2 + \cdots + \frac{c_m \lambda_m^t}{\lambda_1^t} V_m$$

As $|\lambda_i/\lambda_1| < 1$, $(\lambda_i/\lambda_1)^t \to 0$ as $t \to +\infty$. Thus

$$\lim_{t\to+\infty} \frac{X(t)}{\lambda_1^t} = \lim_{t\to+\infty} \frac{L^t X(0)}{\lambda_1^t} = c_1 V_1.$$

Hence after many generations, $X(t) = L^t X(0) = c_1 \lambda_1^t V_1$. The population size either increasing ($\lambda_1 > 1$) or decreasing ($\lambda_1 < 1$) geometrically as t goes larger.

The population distribution $X(t)/\lambda_1^t$ approaches a constant multiple of the eigenvector V_1; thus V_1 is referred to as a stable age distribution. It means that for large values of time, the age distribution vector is a scalar multiple of the eigenvector associated with the largest eigenvalue of the matrix. Consequently the proportion of females in each of the age classes becomes constant, these limiting proportions can be determined from the eigenvector V_1.

An explicit expression for V_1 in the case of a Leslie matrix is [14]

$$V_1 = \begin{pmatrix} 1 \\ \frac{s_1}{\lambda_1} \\ \vdots \\ \frac{s_1 s_2 \ldots s_{m-2}}{\lambda_1^{m-2}} \\ \frac{s_1 s_2 \ldots s_{m-1}}{\lambda_1^{m-1}} \end{pmatrix} \tag{6.6}$$

The characteristic equation for the Leslie matrix satisfies $\det(L - \lambda I) = 0$ or

$$\det \begin{pmatrix} b_1 - \lambda & b_2 & \ldots & b_{m-1} & b_m \\ s_1 & -\lambda & \ldots & 0 & 0 \\ 0 & s_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & s_{m-1} & -\lambda \end{pmatrix} = 0$$

or

$$p(\lambda) = \lambda^m - b_1\lambda^{m-1} - b_2 s_1 \lambda^{m-2} - b_3 s_1 s_2 \lambda^{m-3} - \cdots - b_m s_1 s_2 s_3 \ldots s_{m-1} = 0$$

From Descarte's Rule, since there is only one change in sign in the polynomial, $p(\lambda)$ has one positive real root, that is the dominant eigenvalue λ_1.

How is the dominant eigenvalue λ_1: $\lambda_1 > 1$ or $\lambda_1 < 1$?

- $\lim_{\lambda\to\infty} p(\lambda) = \infty$
- $p(0) < 0$
- $p(\lambda)$ crosses the positive $\lambda-$axis only once at λ_1

then

- $\lambda_1 > 1 \Leftrightarrow p(1) < 0$
- $\lambda_1 < 1 \Leftrightarrow p(1) > 0$

where $p(1) = 1 - b_1 - b_2 s_1 - b_3 s_1 s_2 - \cdots - b_m s_1 s_2 s_3 \ldots s_{m-1}$, and $p(1) = 1 - R_0$. Hence

- $\lambda_1 > 1 \Leftrightarrow 1 < R_0$
- $\lambda_1 < 1 \Leftrightarrow 1 > R_0$

Definition 6.10.2. *The reproductive number R_0 is the average number of offspring produced by an individual in its lifetime:*

$$R_0 = b_1 + b_2 s_1 + b_3 s_1 s_2 + \cdots + b_m s_1 s_2 \ldots s_{m-1}$$

where each term represent the average number of offsprings produced by individuals of age i.

- $R_0 < 1$ *individuals not fully replacing themselves, population shrinking*
- $R_0 = 1$ *individual exactly replacing themselves, population size stable*
- $R_0 > 1$ *individuals more than replacing themselves, population growing*

Theorem 6.10.3. *Assume the Leslie matrix L defined as*

$$X(t+1) = \begin{pmatrix} x_1(t+1) \\ x_2(t+1) \\ x_3(t+1) \\ \vdots \\ x_m(t+1) \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & \ldots & b_{m-1} & b_m \\ s_1 & 0 & \ldots & 0 & 0 \\ 0 & s_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & s_{m-1} & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_m(t) \end{pmatrix} = LX(t)$$

is irreductible and primitive. The characteristic polynomial of L i given by

$$p(\lambda) = \lambda^m - b_1 \lambda^{m-1} - b_2 s_1 \lambda^{m-2} - b_3 s_1 s_2 \lambda^{m-3} - \cdots - b_m s_1 s_2 s_3 \ldots s_{m-1} = 0.$$

Matrix L has a strictly dominant eigenvalue $\lambda_1 > 0$ satisfying the following relationships:

- $\lambda_1 = 1$ *if and only if* $R_0 = 1$,
- $\lambda_1 < 1$ *if and only if* $R_0 < 1$,
- $\lambda_1 > 1$ *if and only if* $R_0 > 1$,

where R_0 is the inherent reproductive number defined by

$$R_0 = b_1 + b_2 s_1 + b_3 s_1 s_2 + \cdots + b_m s_1 s_2 \ldots s_{m-1}.$$

In addition the stable age distribution V_1 satisfies

$$V_1 = \begin{pmatrix} 1 \\ \frac{s_1}{\lambda_1} \\ \vdots \\ \frac{s_1 s_2 \ldots s_{m-2}}{\lambda_1^{m-2}} \\ \frac{s_1 s_2 \ldots s_{m-1}}{\lambda_1^{m-1}} \end{pmatrix}.$$

6.10.1 Salmon population

Suppose a population of salmon live to three years of age. Each adult salmon produces 800 offspring. The probability of a salmon surviving the first year to live on to the second year is 5%, and the probability of a salmon surviving the second year to live on to the third year is 2.5%.

- Find the Leslie matrix for this population.
- If there are 10 females in each of the three age classes, find the initial age distribution vector. Use Matlab to find the population age distribution vectors for each of the first 100 years.
- Use Matlab to find the eigenvalues and eigenvectors of the Leslie Matrix. Is there a strictly dominant eigenvalue?
- Describe what happens to this population of salmon over time?

6.10.2 Human Population

Suppose the population of the United States is broken up into ten 5-year age classes. The values for the reproduction rates F_i and the survival rates P_i for each age class are shown in the table below.

i	F_i	P_i
1	0	0.99670
2	0.00102	0.99837
3	0.08515	0.99780
4	0.30574	0.99672
5	0.40002	0.99607
6	0.28061	0.99472
7	0.15260	0.99240
8	0.06420	0.98867
9	0.01483	0.98274
10	0.00089	0

- Find the Leslie matrix for this population.
- If there are 10 females in each of the ten age classes, find the initial age distribution vector. Use Matlab to find the population age distribution vectors for each of the first 100 years, and plot the age distribution vectors.
- Use Matlab to find the eigenvalues and eigenvectors of the Leslie Matrix. What happens to this population over time?
- After a long period of time, what is the relative number of females in each of the ten age classes?

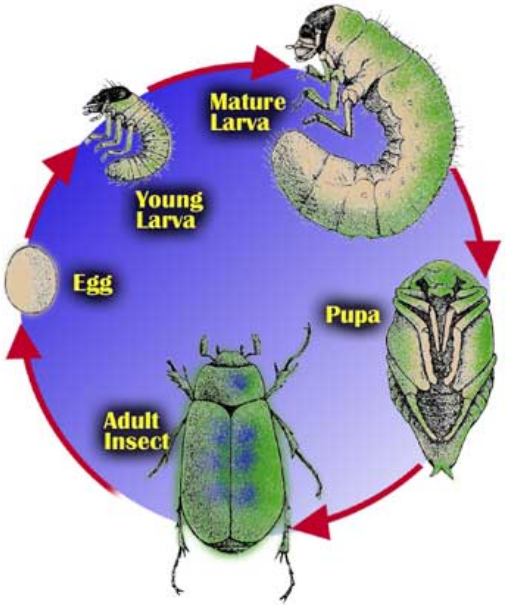


Figure 6.3: Example of insect life cycle.

6.10.3 Insect population

The life cycle of an insect can be decomposed in the following stages.

Adults. The life cycle description can be started with the adult insect. The adult may be a beetle, fly, moth, or midge. Regardless of their form, insects mate and most lay eggs.

Eggs. Eggs come in many shapes, sizes, and colors. They might be deposited on or in the ground, the roots, the stems, the leaves, or the flowers. When the eggs hatch the new insect is called a larva.

Larva. The larva seldom looks like the adult it will become. Some common larval forms are the maggot, grub worm, inchworm, and caterpillar. As the larva grows it must shed it's old skin from time to time. This is called molting. From hatching to the first molt the larva is said to be in it's 1st instar stage. After molting the first time the larva enters it's second instar stage, and so on. The feeding activity of the larvae often inflicts more damage on the noxious weed than the adult form. Different insects have different numbers of instars, but eventually the larva is fully grown and ready to pupate.

Pupa. The pupa is the life stage between larva and adult. In this stage the insect does not feed, and can be considered motionless. This metamorphic change is often profound. Unless the larva is in a stem or root tunnel it will usually construct some kind of shelter to pupate in. This "cocoon" might be made from soil particles, silk, chewed seeds, chewed plant material, ground litter, or combinations. Inside the "cocoon/shelter/chamber/capsule/case" the pupa is gradually transformed into an adult.

6.11 Insect populations

Assume that adult females of a species produce offspring at a fixed period of time each year. A proportion of the offspring (juveniles) survives to adulhood, reproduces, and dies (nonoverlapping of generations). Let

- j_t number of juveniles in years t
- a_t number of adult females in year t
- p number of juveniles that survive in year t
- f number of offspring produced per female
- r ratio of females to adults.

Using the definition of state variables and parameters given above, the model can be expressed as follows

$$a_{t+1} = prj_t$$
$$j_{t+1} = fa_{t+1}$$

The system can be condensed in an unique equation:

$$j_{t+1} = fprj_t$$

If the initial population j_0 of juveniles is known, the solution is unique and is

$$j_t = (fpr)^t j_0 \quad t = 0, 1, 2, \ldots$$

The asymptotic behavior of the solution depends on the value of fpr:

- If $fpr < 1$ $\lim_{t\to+\infty} j_t = 0$ extinction of the population,
- If $fpr > 1$ $\lim_{t\to+\infty} j_t = +\infty$ explosion of the population.

6.12 Pharmacology

A drug is administred once every four hours. Let D_n be the amount of the drug in the blood system at the beginning of the n^{th} interval. The body eliminates a certain fraction p of the drug during each time interval. If the amount administred is D_0, find D_n and $\lim_{n\to\infty} D_n$.

The evolution of the concentration can be described by

$$D_{n+1} = D_n - pD_n + D_0 = (1-p)D_n + D_0 \quad n = 0, 1, \ldots,$$

and the initial condition is D_0. The equilibrium solution is $D^* = \frac{D_0}{p}$ (obtained by solving $D_n = (1-p)D_n + D_0$). Using Proposition 5.2.3 the solution is unique and is

$$D_n = (1-p)^n \left[D_0 - \frac{D_0}{p}\right] + \frac{D_0}{p} \quad n = 0, 1, 2, \ldots$$

The limiting behavior is

$$\lim_{n\to+\infty} D_n = \frac{D_0}{p},$$

since $(1-p) < 1$. Note that if a differential equation formalism were used, the system would take the form

$$\frac{dD}{dt} = -pD + D_0, \quad D(0) = D_0,$$

and the solution would be $D(t) = \frac{D_0}{p} + (D_0 - \frac{D_0}{p})e^{-pt}$.

6.13 Propagation of annual plants

Plants produce seeds at the end of their growth season (August), after which they die. A fraction of these seeds survive the winter, and some of these germinate at the beginning of the season (May), giving rise to the new generation of plants. The fraction that germinates depends on the age of the seeds.

- γ number of seeds produced per plant in August
- σ fraction of seeds that survive a given winter
- α fraction of one-year-old seeds that germinate in May
- β fraction of two-year-old seeds that germinate in May
- Seeds older than two years are no longer viable

State variables:

- p_n number of plants in generation n
- s_n number of new seed in generation n
- s_n^1 number of one-year-old seeds in generation n
- s_n^2 number of two-year-old seeds in generation n

Equations:

$$p_n = \alpha s_n^1 + \beta s_n^2 \tag{6.7a}$$
$$s_n = \gamma p_n \tag{6.7b}$$
$$s_n^1 = \sigma s_{n-1} \tag{6.7c}$$
$$s_n^2 = \sigma(1-\alpha) s_{n-1}^1 \tag{6.7d}$$

Condensing equations (6.7b), (6.7c) and (6.7d) we obtain

$$s_n^1 = \sigma\gamma p_{n-1}$$

and

$$s_n^2 = \sigma(1-\alpha)\sigma\gamma p_{n-2}$$

therefore, we can express the model as a system of 3 First-order difference equations

$$\begin{aligned} p_n &= \alpha s_n + \beta\sigma(1-\alpha)s_{n-1} \\ s_n^1 &= \sigma\gamma p_{n-1} \\ s_n^2 &= \sigma(1-\alpha)\sigma\gamma p_{n-2} \end{aligned}$$

or as 1 Second-order equation:

$$p_n = \alpha\sigma\gamma p_{n-1} + \beta\sigma(1-\alpha)\sigma\gamma p_{n-2} \tag{6.8}$$

The characteristic equation corresponding to the model (6.8) is

$$\lambda^2 - \alpha\sigma\gamma\lambda - \beta\sigma^2(1-\alpha)\gamma = 0.$$

Eigenvalues are

$$\lambda_{1,2} = \frac{\alpha\sigma\gamma \pm \sqrt{(\alpha\sigma\gamma)^2 + 4\beta\sigma^2(1-\alpha)\gamma}}{2}$$

and

$$\lambda_{1,2} = \frac{\alpha\sigma\gamma}{2}\left(1 \pm \sqrt{1 + \frac{4\beta(1-\alpha)}{\alpha^2\gamma}}\right).$$

The dominant eigenvalue (the eigenvalue corresponding to the solution that determines the limiting behavior of the general solution) is the positive eigenvalue

$$\lambda_1 = \frac{\alpha\sigma\gamma}{2}\left(1 + \sqrt{1 + \frac{4\beta(1-\alpha)}{\alpha^2\gamma}}\right).$$

If $0 < \lambda_1 < 1$ the plant population will become extinct. If $\lambda_1 > 1$ the plant population will grow.

6.14 Red blood cells

In the circulatory system, red blood cells are constantly being destroyed and replaced. They carry oxygen throughout the body and they must be maintained at a constant level. The spleen filters out and destroys a fraction of the cells daily and the bone marrow produces a number proportional to the number lost on the previous day. The cell count on dat t is modeled as followed:

- R_t number of red blood cells in circulation on day t.
- M_t number of red blood cells produced by marrow on day t.
- f fraction of red blood cells removed by spleen, $0 < f < 1$.
- γ production constant, $\gamma > 0$

The system of difference equations is

$$R_{t+1} = (1-f)R_t + M_t \tag{6.9a}$$
$$M_{t+1} = \gamma f R_t \tag{6.9b}$$

or a Second-order difference equation:

$$R_{t+1} = (1-f)R_t + \gamma f R_{t-1} \tag{6.10}$$

Characteristic equation corresponding to the model (6.10) is

$$\lambda^2 - (1-f)\lambda - \gamma f = 0$$

Eigenvalues are

$$\lambda_{1,2} = \frac{(1-f) \pm \sqrt{(1-f)^2 + 4\gamma f}}{2}$$

For homeostasis in the red blood cell count, the total number of red blood cell has to stay constant. Then the dominanting eigenvalue has to be $\lambda_1 = 1$.

The dominating eigenvalue is

$$\lambda_1 = \frac{(1-f) + \sqrt{(1-f)^2 + 4\gamma f}}{2}.$$

By investigating $\lambda_1 = 1$, we found that the condition, $\gamma = 1$, has to be satisfied.

If $\gamma = 1$, $\lambda_2 = -f$, and the solution is

$$R_t = c_1(-f^t) + c_2$$

the system is oscillating but decreasing in magnitude to reach a constant number.

6.15 Killer whales

Killer whales are long-lived marine mammals that live in stable social groups called "pods". Demographic data on killer whale populations in the coastal waters of British Columbia and Washington state have been collected since 1973. Brault and Caswell (1993) used the 1973-1987 data and a stage-structured matrix model to investigated several demographic questions concerning the whales. They model the females with a mixed age-stage classification: yearlings, juveniles (past the first year, but not mature), mature females, and post-reproductive females.

$$A = \begin{pmatrix} 0 & 0.0043 & 0.1132 & 0 \\ 0.9775 & 0.9111 & 0 & 0 \\ 0 & 0.0736 & 0.9534 & 0 \\ 0 & 0 & 0.0452 & 0.9804 \end{pmatrix}$$

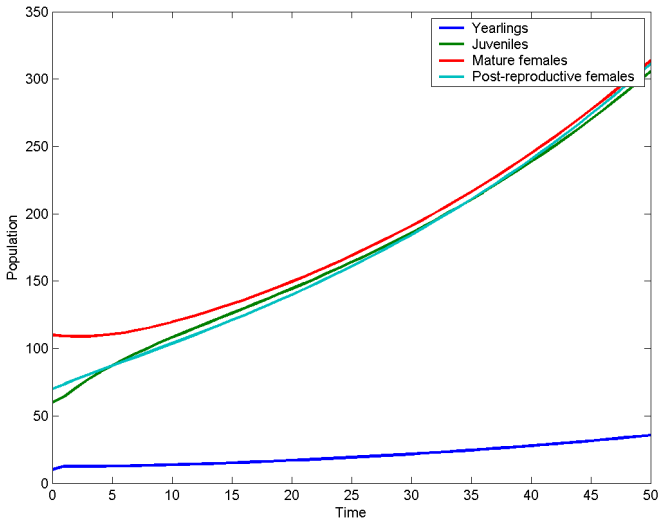
- Draw the life-cycle graph associated with the matrix A.
- Computes the dominant eigenvalue.
- Find the stable stage distribution for the whale population.
- Projects the population dynamics for the next 50 years assuming that the current population vector is $x_0 = (10, 60, 110, 70)$.

MatLab code:

```
L=[0 0.0043 0.1123 0;.9775 0.9111 0 0;0 .0736 0.9534 0;0 0 0.0452 0.9804];
x0=[10;60;110;70];
X=zeros(4,51);
X(:,1)=x0;
```

```
%simulations
for k=2:51, X(:,k)=L*X(:,k-1); end
t=0:50;
plot(t,X);
xlabel('Time');
ylabel('Population');
legend('Yearlings', 'Juveniles', 'Mature females','Post-reproductive females')
% limiting behavior
L=[0 0.0043 0.1123 0;.9775 0.9111 0 0;0 .0736 0.9534 0;0 0 0.0452 0.9804];
Value=eig(L);
dominant=max(Value)
[V,D]=eig(L)

dominant =
    1.0251

V =
         0    0.0658   -0.0655    0.6788
         0    0.5640    0.8371   -0.7321
         0    0.5789   -0.5187    0.0568
    1.0000    0.5852    0.1609   -0.0026

D =
    0.9804         0         0         0
         0    1.0251         0         0
         0         0    0.8346         0
         0         0         0    0.0048

>> v1=V(:,2)/sum(V(:,2))

v1 =
    0.0367
    0.3144
    0.3227
    0.3262
```

The dominant eigenvalue is 1.0251, and the associated eigenvector is $(0.0658, 0.5640, 0.5789, 0.5852)^T$. By summing all the entries of this eigenvector and by dividing each of its entries by this sum, we can express the stable stage distribution v_1 for the whale population.


Yearlings
Juveniles
Mature females
Post-reproductive females
Population
Time
0
50
100
150
200
250
300
350
0
5
10
15
20
25
30
35
40
45
50

Part III

Markov chains

Chapter 7

A brief theory of Markov chains

7.1 Markov chains

We conduct an experiment with a set of r *outcomes* or *states*,

$$S = \{S_1, \ldots, S_r\}.$$

The experiment is repeated n times (with n large, potentially infinite). The system has no memory: the next state depends only on the present state. The probability of S_j occurring on the next step, given that S_i occurred on the last step, is

$$p_{ij} = p(S_j|S_i).$$

Suppose that S_i is the current state, then one of $S_1, \ldots, S_r$ must be the next state; therefore,

$$p_{i1} + p_{i2} + \cdots + p_{ir} = 1, \quad 1 \le i \le r.$$

(Note that some of the p_{ij} can be zero, all that is needed is that $\sum_{j=1}^{r} p_{ij} = 1$ for all i.)

Definition 7.1.1. *An experiment with finite number of possible outcomes $S_1, \ldots, S_r$ is repeated. The sequence of outcomes is a* Markov chain *if there is a set of r^2 numbers $\{p_{ij}\}$ such that the conditional probability of outcome S_j on any experiment given outcome S_i on the previous experiment is p_{ij}, i.e., for $1 \le i, j \le r$, $n = 1, \ldots$,*

$$p_{ij} = \mathsf{Pr}(S_j \text{ on experiment } n+1 | S_i \text{ on experiment } n).$$

The outcomes $S_1, \ldots, S_r$ are the states*, and the p_{ij} are the* transition probabilities*. The matrix $P = [p_{ij}]$,*

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix},$$

is called the transition matrix *of the Markov chain.*

The transition matrix has

- nonnegative entries, $p_{ij} \ge 0$

- entries less than 1, $p_{ij} \leq 1$
- row sum equal to 1, which we write

$$\sum_{j=1}^{r} p_{ij} = 1, \quad i = 1, \ldots, r$$

or, using the notation $\mathbb{1}^T = (1, \ldots, 1)$,

$$P\mathbb{1} = \mathbb{1}.$$

These properties guarantee that the elements of the matrix P describe probabilities.

7.2 Repetition of the process

Let $p_i(n)$ be the probability that the state S_i will occur on the n^{th} repetition of the experiment, $1 \leq i \leq r$. Since one the states S_i must occur on the n^{th} repetition,

$$p_1(n) + p_2(n) + \cdots + p_r(n) = 1.$$

Let $p_i(n+1)$ be the probability that state S_i, $1 \leq i \leq r$, occurs on $(n+1)^{th}$ repetition of the experiment. There are r ways to be in state S_i at step $n+1$:

1. Step n is S_1. Probability of getting S_1 on n^{th} step is $p_1(n)$, and probability of having S_i after S_1 is p_{1i}. Therefore, $P(S_i|S_1) = p_{1i}p_1(n)$.
2. We get S_2 on step n and S_i on step $(n+1)$. Then $P(S_i|S_2) = p_{2i}p_2(n)$.

..

r. Probability of occurrence of S_i at step $n+1$ if S_r at step n is $P(S_i|S_r) = p_{ri}p_r(n)$.

Therefore, $p_i(n+1)$ is sum of all these probabilities,

$$\begin{aligned} p_i(n+1) &= P(S_i|S_1) + \cdots + P(S_i|S_r) \\ &= p_{1i}p_1(n) + \cdots + p_{ri}p_r(n) \end{aligned}$$

Therefore,

$$\begin{aligned} p_1(n+1) &= p_{11}p_1(n) + p_{21}p_2(n) + \cdots + p_{r1}p_r(n) \\ &\vdots \\ p_k(n+1) &= p_{1k}p_1(n) + p_{2k}p_2(n) + \cdots + p_{rk}p_r(n) \\ &\vdots \\ p_r(n+1) &= p_{1r}p_1(n) + p_{2r}p_2(n) + \cdots + p_{rr}p_r(n) \end{aligned} \tag{7.1}$$

In matrix form

$$p(n+1) = p(n)P, \quad n = 1, 2, 3, \ldots \tag{7.2}$$

where $p(n) = (p_1(n), p_2(n), \ldots, p_r(n))$ is a (row) probability vector and $P = (p_{ij})$ is a $r \times r$ *transition matrix*,

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

So, what we have is

$$(p_1(n+1), \ldots, p_r(n+1)) = (p_1(n), \ldots, p_r(n)) \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

It is easy to check that this gives the same expression as (7.1).

7.2.1 Long time behaviour

Let $p(0)$ be the initial distribution (row) vector. Then

$$\begin{aligned} p(1) &= p(0)P \\ p(2) &= p(1)P \\ &= (p(0)P)P \\ &= p(0)P^2 \end{aligned}$$

Iterating, we get that for any n,

$$p(n) = p(0)P^n$$

Therefore,

$$\lim_{n\to+\infty} p(n) = \lim_{n\to+\infty} p(0)P^n = p(0) \lim_{n\to+\infty} P^n. \tag{7.3}$$

So, to determine the long time behaviour of the Markov chain, it suffices to determine the limit of the P^n.

7.2.2 Stochastic matrices

One fundamental property of Markov chains is that the transition matrix P has a very particular structure.

Definition 7.2.1 (Stochastic matrix)**.** *The nonnegative $r \times r$ matrix M is a* stochastic *matrix if $\sum_{j=1}^{r} a_{ij} = 1$ for all $i = 1, 2, \ldots, r$, i.e., the rows of M all sum to 1.*

Note that a matrix that has column sums all equal to 1 is also called a stochastic matrix. Such a matrix arises if instead of representing transitions from rows to columns as we have done, transitions are represented from columns to rows. If it is needed to distinguish between matrices that are stochastic by rows or by columns, we say that, say, a matrix is a (row) stochastic matrix.

A matrix that that both rows and column sums equal to 1 is a *doubly stochastic* matrix. Stochastic matrices over $\mathbb{R}_+$ form a *group*. This means in particular that the following holds true.

Theorem 7.2.2. *If M, N are stochastic matrices, then MN is a stochastic matrix.*

This is particularly interesting, as it means that powers of a stochastic matrix are also stochastic:

Theorem 7.2.3. *If M is a stochastic matrix, then for any $k \in \mathbb{N}$, M^k is a stochastic matrix.*

A stochastic matrix has the property that all its eigenvalues are contained in the unit disk of $\mathbb{C}$.

Theorem 7.2.4. *Let M be a stochastic matrix M. Then the following two properties hold.*

i) The spectral radius *of M, $\rho(M)$, is such that $\rho(M) = 1$. In other words, all eigenvalues λ of M are such that $|\lambda| \leq 1$. Furthermore, $\lambda = 1$ is an eigenvalue of M.*

ii) If M is stochastic by rows, then $\lambda = 1$ is associated to the eigenvector $\mathbb{1}$, while if M is stochastic by columns, then $\lambda = 1$ is associated to the left eigenvector $\mathbb{1}^T$.

Proof. To see that 1 is an eigenvalue, write the definition of a stochastic matrix, i.e., M has row sums 1. In vector form, $M\mathbb{1} = \mathbb{1}$. Now remember that λ is an eigenvalue of M, with associated eigenvector v, iff $Mv = \lambda v$. So, in the expression $M\mathbb{1} = \mathbb{1}$, we read an eigenvector, $\mathbb{1}$, and an eigenvalue, 1. □

7.3 Regular Markov chains

Definition 7.3.1 (Regular Markov chain)**.** *A regular Markov chain is one in which P^k is positive for some integer $k > 0$, i.e., P^k has only positive entries, no zero entries.*

In matrix theory, this property is called *primitivity.*

Definition 7.3.2. *A nonnegative matrix M is primitive if, and only if, there is an integer $k > 0$ such that M^k is positive.*

Therefore, we have the following result.

Theorem 7.3.3. *A Markov chain is regular if, and only if, the transition matrix P is primitive.*

Another way to check regularity will then be using a sufficient condition for primitivity, given here.

Theorem 7.3.4. *A matrix M is primitive if the associated connection graph is strongly connected, i.e., that there is a path between any pair (i, j) of states, and that there is at least one positive entry on the diagonal of M.*

This is checked directly on the digraph associated to the transition matrix, shown for our model in Figure 8.1. The digraph is here strongly connected, and as there are positive entries on the diagonal of the matrix (as evidenced by the self-connecting loops on all vertices), the transition matrix associated to this graph is primitive, from Theorem 7.3.4.

Theorem 7.3.5. *If P is the transition matrix of a regular Markov chain, then*

i) the powers P^n approach a stochastic matrix W,

ii) each row of W is the same (row) vector $w = (w_1, \dots, w_r)$,

iii) the components of w are positive.

So if the Markov chain is regular,

$$\lim_{n\to+\infty} p(n) = p(0) \lim_{n\to+\infty} P^n = p(0)W. \tag{7.4}$$

Application to Markov chains We already know that the (right) eigenvector corresponding to the eigenvalue 1 is $\mathbb{1}$. The vector w in Theorem 7.3.5 is in fact the left eigenvector corresponding to the eigenvalue 1 of P. To see this, remark that, if $p(n)$ converges, then $p(n+1) = p(n)P$, so w is a fixed point of the system. We thus write

$$wP = w$$

and solve for w, which amounts to finding w as the left eigenvector corresponding to the eigenvalue 1.

Alternatively, we can find w as the (right) eigenvector associated to the eigenvalue 1 for the transpose of P,

$$P^T w^T = w^T$$

Now remember that when you compute an eigenvector, you get a result that is the eigenvector, to a multiple. So the expression you obtain for w might have to be normalized (you want a probability vector). Once you obtain w, check that the norm $\|w\|$ defined by

$$\|w\| = w_1 + \cdots + w_r$$

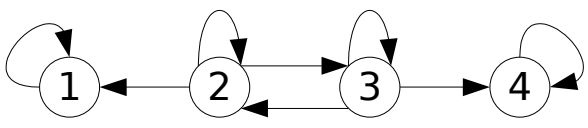
is equal to one. If not, use

$$\frac{w}{\|w\|}.$$

Definition 7.3.6. *A state S_i in a Markov chain is* absorbing *if whenever it occurs on the n^{th} generation of the experiment, it then occurs on every subsequent step. In other words, S_i is absorbing if $p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$. A state that is not absorbing is called* transient.

Definition 7.3.7. *A Markov chain is said to be absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state.*

Suppose we have a chain like the following:



i) Does the process eventually reach an absorbing state?

ii) Average number of times spent in a transient state, if starting in a transient state?

iii) Average number of steps before entering an absorbing state?

iv) Probability of being absorbed by a given absorbing state, when there are more than one, when starting in a given transient state?

Answer to question 1:

Theorem 7.3.8. *In an absorbing Markov chain, the probability of reaching an absorbing state is 1.*

For an absorbing chain with k absorbing states and $r-k$ transient states, the transition matrix can be written in the following **standard form**:

$$P = \begin{pmatrix} \mathbb{I}_k & \mathbf{0} \\ R & Q \end{pmatrix}$$

with following meaning,

	Absorbing states	Transient states
Absorbing states	$\mathbb{I}_k$	$\mathbf{0}$
Transient states	R	Q

with $\mathbb{I}_k$ the $k \times k$ identity matrix, $\mathbf{0}$ an $k \times (r-k)$ matrix of zeros, R an $(r-k) \times k$ matrix and Q an $(r-k) \times (r-k)$ matrix.

The matrix $\mathbb{I}_{r-k} - Q$ is invertible. Let

- $N = (\mathbb{I}_{r-k} - Q)^{-1}$ be the *fundamental matrix* of the Markov chain
- T_i be the sum of the entries on row i of N
- $B = NR$.

Answers to our remaining questions:

ii) N_{ij} is the average number of times the process is in the jth transient state if it starts in the ith transient state.

iii) T_i is the average number of steps before the process enters an absorbing state if it starts in the ith transient state.

iv) B_{ij} is the probability of eventually entering the jth absorbing state if the process starts in the ith transient state.

Definition 7.3.9. *A regular Markov chain is one in which S^p is positive for some positive integer p.*

From Theorem B.2.2

$$\rho(S) \leq \|S\|_1 = 1$$

then $|\lambda| \leq 1$ for all eigenvalues of a stochastic matrix. Furthermore, if S is a stochastic matrix $\lambda = 1$ is an eigenvalue of S; hence $\rho(S) = 1$ and the dominant eigenvalue $\lambda_1 = 1$. Then

$$\lim_{n \to +\infty} p(n) = \lim_{n \to +\infty} S^n p(0) = cV_1$$

where $V_1 = (v_1, v_2, \ldots, v_k)$ is the eigenvector that corresponds to the dominant eigenvalue $\lambda_1 = 1$.

Since $p(n) = (p_1(n), p_2(n), \ldots, p_k(n))^T$, we have $\sum_{i=1}^{k} p_i(n) = 1$, it follows that

$$cv_1 + cv_2 + \ldots cv_k = 1.$$

Therefore

$$c = \frac{1}{v_1 + v_2 + \cdots + v_k}.$$

Definition 7.3.10. *A state s_i in a Markov chain is said to be absorbing if whenever it occurs on the n^{th} generation of the experiment, it then occurs on every subsequent repetition. In other word, if for some $p_{ii} = 1$ then $p_{ij} = 0$ for $i \neq j$.*

Definition 7.3.11. *A Markov chain is said to be absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state.*

In an absorbing Markov chain, a state that is not absorbing is called transient.

Chapter 8

Models using Markov chains

8.1 A simple genetic model

The simplest type of genetic inheritance of traits in animals occurs when a certain trait is determined by a specific pair of genes, each of which may be two types, say G and g. An individual may have a GG combination, a Gg (genetically equivalent to gG), or gg combination. An individual with GG is said to be dominant, a gg individual is recessive and a Gg is an hybrid.

In the mating of two animals, the offspring inherits one gene of the pair from each parent: the basic assumption of genetics is that these genes are selected at random, independently of each other.

This assumption determines the probability of occurrence of each type of offspring: The offspring

- of two dominant parents must be dominant,
- of two recessive parents must be recessive,
- and of one dominant and one recessive parent must be hybrid.

In the mating of a dominant and a hybrid animal, each offspring must get a G gene from the former and has an equal chance of getting G or g from the latter. Hence there is an equal probability of getting a dominant or a hybrid offspring. Again, in the mating of a recessive and a hybrid, there is an even chance for getting either a recessive or a hybrid. In the mating of two hybrids, the offspring has an equal chance of getting G or g from each parent. Hence the probabilities are $1/4$ for GG, $1/2$ for Gg, and $1/4$ for gg.

A certain trait is determined by a specific pair of genes, each of which may be two types, say G and g. One individual may have:

- GG combination (*dominant*)
- Gg or gG, considered equivalent genetically (*hybrid*)
- gg combination (*recessive*)

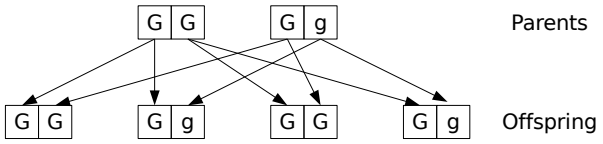
In sexual reproduction, offspring inherit one gene of the pair from each parent.

8.1.1 Basic assumption of Mendelian genetics

Genes inherited from each parent are selected at random, independently of each other. This determines probability of occurrence of each type of offspring. The offspring

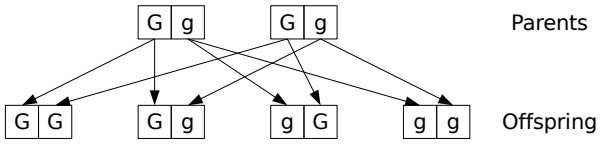
- of two GG parents must be GG,
- of two gg parents must be gg,
- of one GG and one gg parent must be Gg,
- other cases must be examined in more detail.

GG and Gg parents



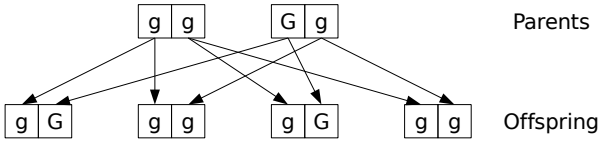
Offspring has probability $\frac{1}{2}$ of being GG and $\frac{1}{2}$ of being Gg.

Gg and Gg parents



Offspring has probability $\frac{1}{4}$ of being GG, $\frac{1}{2}$ of being Gg and $\frac{1}{4}$ of being gg.

gg and Gg parents



Offspring has probability $\frac{1}{2}$ of being Gg and $\frac{1}{2}$ of being gg.

8.1.2 A first genetic model – Regular Markov chain

Consider a process of continued matings with the following characteristics:

- Start with an individual of known or unknown genetic character and mate it with a hybrid.
- Assume that there is at least one offspring; choose one of them at random and mate it with a hybrid.

- Repeat this process through a number of generations.

The genetic type of the chosen offspring in successive generations can be represented by a Markov chain, with states GG, Gg and gg. So there are 3 possibles states $S_1 = GG$, $S_2 = Gg$ and $S_3 = gg$. We have

$\nearrow$	GG	Gg	gg
GG	0.5	0.5	0
Gg	0.25	0.5	0.25
gg	0	0.5	0.5

The transition probabilities are thus

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

The (di)graph associated to P is shown in Figure 8.1. The Markov chain with transition matrix

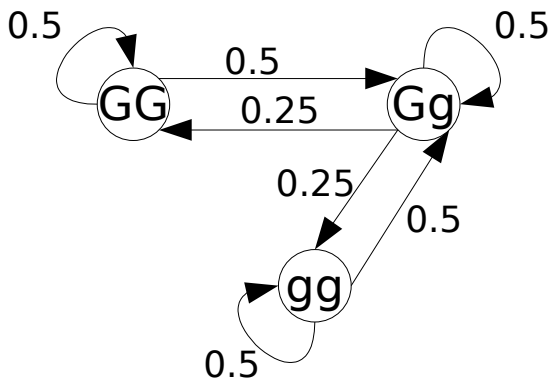


Figure 8.1: Transition graph for the Markov chain for the genetics model with mating with an hybrid individual.

P is regular. Indeed, compute P^2:

$$P^2 = \begin{pmatrix} \frac{3}{8} & \frac{1}{2} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \end{pmatrix}$$

As all entries are positive, P is primitive and the Markov chain is regular. It is also possible to check this information directly by looking at the graph associated to the Markov chain (Figure 8.1). There, it is easy to check that it is possible to go from any state to any other state, i.e., the graph is strongly connected. From Theorem **??**, it follows that the matrix P is irreducible. Further, the self-connecting loops mean that Theorem **??** can be used, implying the primitivity of P.

Compute the left eigenvector associated to 1 by solving

$$(w_1, w_2, w_3) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (w_1, w_2, w_3)$$

$$\frac{1}{2}w_1 + \frac{1}{4}w_2 = w_1 \tag{8.1a}$$

$$\frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_3 = w_2 \tag{8.1b}$$

$$\frac{1}{4}w_2 + \frac{1}{2}w_3 = w_3 \tag{8.1c}$$

From (8.1a), $w_1 = w_2/2$, and from (8.1c), $w_3 = w_2/2$. Substituting these values into (8.1b),

$$\frac{1}{4}w_2 + \frac{1}{2}w_2 + \frac{1}{4}w_2 = w_2,$$

that is, $w_2 = w_2$, i.e., w_2 can take any value. So $w = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

8.1.3 A second genetic model – Absorbing Markov chain

Suppose now that we conduct the same experiment, but mate the individual picked at random in each new generation with a GG individual instead of a Gg individual. The transition table is

$\nearrow$	GG	Gg	gg
GG	1	0	0
Gg	0.5	0.5	0
gg	0	1	0

and the resulting transition probabilities are

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

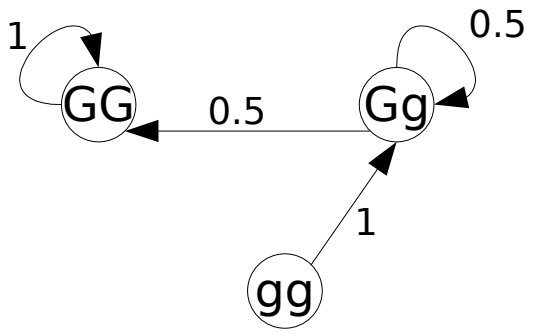


Figure 8.2: Transition digraph for the Markov chain for the genetics model with mating with a dominant individual. This graph is not strongly connected; vertex GG is a *sink*, while vertex gg is a *source*, or, in Markov chain vocabulary, state GG is absorbing and state gg is repulsive.

Clearly,

- the process leaves gg after one iteration, and can never return,
- as soon as the process leaves Gg, it can never return there,
- and it can never leave GG as soon as it gets there.

The matrix is already in standard form,

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \mathbb{I}_1 & \mathbf{0} \\ R & Q \end{pmatrix},$$

with $\mathbb{I}_1 = 1$, $\mathbf{0} = (0 \;\; 0)$ and

$$R = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} \qquad Q = \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{pmatrix}.$$

We have

$$\mathbb{I}_2 - Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ -1 & 1 \end{pmatrix},$$

so

$$N = (\mathbb{I}_2 - Q)^{-1} = 2 \begin{pmatrix} 1 & 0 \\ 1 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}.$$

Then

$$T = N\mathbb{1} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and

$$B = NR = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Part IV

Ordinary differential equations

Chapter 9

A brief theory of ordinary differential equations

In this chapter, we consider ordinary differential equations of first-order,

$$x' = f(t, x), \tag{9.1}$$

with $x \in \mathbb{R}^n$ and $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, as well as pth-order ordinary differential equations of the form

$$x^{(p)} = f(t, x, x', \ldots, x^{(p-1)}), \tag{9.2}$$

where $x \in \mathbb{R}$ and $x^{(k)}$ denotes the kth derivative of $x(t)$. These two equations are said to be in *normal form*, and form a subset of a larger class of first-order and pth-order differential equations, defined respectively by $F(t, x, x') = 0$ and $F(t, x, x', \ldots, x^{(p)}) = 0$. We could also consider vector-valued pth-order equations. However, (9.3) and (9.2) are sufficient for our purpose, and we limit our attention to them.

9.1 First definitions

A system such as (9.3) is *nonautonomous*, since the function f depends explicitly on t (time). Much of the theory in this chapter will concern specifically the *autonomous* system,

$$x' = f(x), \tag{9.3}$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$.

Definition 9.1.1. *An equilibrium solution of equation* (9.3) *is a solution* $\bar{x}$ *satisfying*

$$f(\bar{x}) = 0.$$

The system at $x = \bar{x}$ is then said to be at equilibrium. Indeed, since $f(\bar{x}) = 0$, we have

$$\frac{d}{dt}\bar{x}(t) = f(\bar{x}(t)) = 0,$$

and thus the system is "at rest". Equilibria are then classified in terms of their *stability*. We return to this concept later, but we give the definitions now because they appear throughout this chapter.

Definition 9.1.2. *(Local stability) An equilibrium solution $\bar{x}$ of* (9.3) *is* locally stable *if for each $\epsilon > 0$ there exits a $\delta > 0$ such that every solution $x(t)$ of* (9.3) *with the initial condition $x(t_0) = x_0$,*

$$\|x_0 - \bar{x}\|_2 < \delta,$$

satisfies the condition that

$$\|x(t) - \bar{x}\|_2 < \epsilon$$

for all $t \geq t_0$. If the equilibrium solution is not locally stable it is said to be unstable.

Euclidian distance between two points $Y_1 = (y_1^1, y_2^1, \ldots, y_n^1)$ and $Y_2 = (y_1^2, y_2^2, \ldots, y_n^2)$ in $\mathbb{R}^n$ is

$$\|Y_1 - Y_2\|_2 = \sqrt{\sum_{i=1}^{n} (y_i^1 - y_i^2)^2}.$$

Definition 9.1.3. *(Local asymptotic stability) An equilibrium solution $\bar{x}$ of* (9.3) *is* locally asymptotically stable *if it is locally stable and if there exist $\gamma > 0$ such that $\|x_0 - \bar{x}\|_2 < \gamma$ implies*

$$\lim_{t \to \infty} \|x(t) - \bar{x}\|_2 = 0.$$

Definition 9.1.4. *(Periodic solution) A* periodic solution *of the system* (9.3) *is a nonconstant solution $x(t)$ satisfying $x(t + T) = x(t)$ for all t on the interval of existence for some $T > 0$. The minimum value of T is called the period of the solution.*

9.2 First-order differential equations

Definition 9.2.1. *The standard form of first order linear equations is*

$$\frac{dy}{dt} + p(t)y = g(t)$$

p and g are given functions of the independent variable t.

9.2.1 Analytical methods

Linear equations: Integrating factors

To solve 1^{st} order linear equation with non-constant coefficients (this method can also be used for equation with constant coefficient).

i) Put the DE in the standard form

$$\frac{dy}{dt} + p(t)y = g(t) \tag{9.4}$$

ii) Determine the integrating factor $\mu(t)$

- Multiply the DE (9.4) by $\mu(t)$

$$\mu(t)\frac{dy}{dt} + \mu(t)p(t)y = \mu(t)g(t) \tag{9.5}$$

- State that the left side of (9.5) is equal to $\frac{d}{dt}(\mu(t)y)$

$$\frac{d}{dt}(\mu(t)y) = \mu(t)\frac{dy}{dt} + y\frac{d\mu}{dt} = \mu(t)\frac{dy}{dt} + \mu(t)p(t)y$$

- Solve for $\mu(t)$

$$\frac{d\mu}{dt} = \mu(t)p(t)$$

$$\Rightarrow \quad \mu(t) = e^{\int p(t)dt}$$

iii) Solve (9.5) for y with $\mu(t) = e^{\int p(t)dt}$

$$\left(\frac{d}{dt}e^{\int p(t)dt}y = \right) e^{\int p(t)dt}\frac{dy}{dt} + p(t)e^{\int p(t)dt}y = e^{\int p(t)dt}g(t)$$

$$\frac{d}{dt}\mu(t)y = \mu(t)g(t)$$

$$\mu(t)y = \int \mu(t)g(t)dt + c$$

Hence the general solution of (9.4) is

$$y(t) = \frac{1}{\mu(t)}\left[\int_{t_0}^{t} \mu(s)g(s)ds + c\right] \quad with \quad \mu(t) = e^{\int p(t)dt}$$

Separable equations

Definition 9.2.2. *(Separable equations) A first order differential equation*

$$\frac{dy}{dx} = f(x, y)$$

is said to be separable or to have separable variables if it can be expressed as follows

$$\frac{dy}{dx} = g(x)h(y)$$

(the rate function can be expressed as a product of a function of the independent variable times a function of the dependent variable).

To solve separable equations: $\frac{dy}{dx} = g(x)h(y)$

i) Express the separable equation as follows

$$\frac{1}{h(y)}\frac{dy}{dx} = g(x)$$

ii) As y, $\frac{dy}{dx}$, and $g(x)$ are functions of x, apply the integral

$$\int \frac{1}{h(y)}\frac{dy}{dx}dx = \int g(x)dx$$

iii) Use the Change of variable Theorem (if $u = v(x)$ $\int f(v(x))v'(x)dx = \int f(u)du$) for the left side with $u = y(x)$

$$\int \frac{1}{h(u)}du = \int g(x)dx$$
$$\int \frac{1}{h(y)}dy = \int g(x)dx$$

iv) Integrate

$$H(y) = G(x) + c \tag{9.6}$$

c is the combination of the left and right integration constants, H and G are antiderivatives of $\frac{1}{h(y)}$ and $g(x)$ respectively.

v) Solve equation (9.6) for y to obtain a explicit form of the general solution.

Local stability

A simple criterion for determining the L.A.S. of an equilibrium solution to a first order autonomous differential equation

$$\frac{dy}{dt} = f(y) \tag{9.7}$$

having an equilibrium at $\bar{y}$.

Theorem 9.2.3. *Suppose f' is continuous on an open interval I containing $\bar{y}$, where $\bar{y}$ is an equilibrium of $\frac{dy}{dt} = f(y)$. Then $\bar{y}$ is locally asymptotically stable if*

$$f'(\bar{y}) < 0$$

and unstable if $f'(\bar{y}) > 0$.

The value $f'(\bar{y})$ is know as the eigenvalue of the linearized equation.

Definition 9.2.4. *The equilibrium $\bar{y}$ of $\frac{dy}{dt} = f(y)$ is called hyperbolic if $f'(\bar{y}) \neq 0$. Otherwise, it is called nonhyperbolic.*

Nonhyperbolic equilibria have a zero eigenvalue, and hence their local stability is indeterminate.

Phase line analysis

Consider a first-order nonautonomous equation

$$\frac{dy}{dt} = f(t, y).$$

$f(t, y)$ represents the slope of the tangent to the solution $y(t)$ at the point (t, y). Hence $f(t, y)$ can be used to construct a direction field in the $t, y-$plane.

Consider a first-order autonomous equation

$$\frac{dy}{dt} = f(y).$$

Here the direction of flow does not change with t. Hence it is only necessary to determine the direction of flow on the $y-$axis (Phase line). If $\frac{dy}{dt}$ is positive, the direction of flow is in the positive direction, if it is negative, the flow is in negative direction.

9.2.2 Higher-order linear equations

The general solution to an $n^{th}-$order linear nonhomogeneous differential equation is the sum of two solutions, a general solution to the homogeneous differential equation and a particular solution to the nonhomogeneous differential equation

$$y(t) = y_h(t) + y_p(t).$$

• Methodology to find the general solution to the homogeneous equation is given for second-order linear equations, but it can be generalized to the $n^{th}-$order linear equations.

Theorem 9.2.5. *(Principle of superposition) If y_1 and y_2 are two solutions of the homogeneous linear differential equation on an interval I*

$$y'' + p(t)y' + q(t)y = 0$$

then the linear combination

$$c_1y_1 + c_2y_2$$

is also a solution for any values of the constants c_1 and c_2 on the interval.

Definition 9.2.6. *(Wronskian) Suppose each of the functions $f_1(t), f_2(t), \ldots, f_n(t)$ has at least $n-1$ derivatives. The determinant*

$$W(f_1, f_2, \ldots, f_n) = \begin{vmatrix} f_1 & f_2 & \cdots & f_n \\ f_1' & f_2' & \cdots & f_n' \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{vmatrix}$$

is called the Wronskian of the functions.

Theorem 9.2.7. *Let y_1 and y_2 be solutions of the differential equation*

$$y'' + p(t)y' + q(t)y = 0$$

where p and q are continuous on an open interval I. Then y_1 and y_2 are linearly independent on I if and only if $W(y_1, y_2)(t) \neq 0\ \forall t \in I$.

Definition 9.2.8. *Any set y_1, y_2 of 2 linearly independent solutions of*

$$y'' + p(t)y' + q(t)y = 0$$

on an open interval I is said to be **a fundamental set of solutions** *on I of the differential equation.*

Definition 9.2.9. *The general solution of*

$$y'' + p(t)y' + q(t)y = 0$$

on an open interval I is the linear combination of 2 linearly independent solutions y_1 and y_2

$$y(t) = c_1y_1(t) + c_2y_2(t)$$

with c_1 and c_2 constants. y_1 and y_2 form a fundamental set of solutions.

Method for linear homogeneous second-order equations with constant coefficients

$$ay'' + by' + cy = 0 \tag{9.8}$$

Let assume that the solution can take the form of

$$y(t) = e^{rt}.$$

i) Write the **characteristic equation**

$$ar^2 + br + c = 0$$

ii) Find the **roots** r_1 and r_2 of the characteristic equation

$$r_1 = \frac{-b - \sqrt{\Delta}}{2a}, \qquad r_2 = \frac{-b + \sqrt{\Delta}}{2a}, \quad \text{with} \quad \Delta = b^2 - 4ac$$

$\Delta > 0$: r_1, r_2 distinct and real, $y_1(t) = e^{r_1 t}$ and $y_2(t) = e^{r_2 t}$ are 2 linearly independent solutions of (9.8). The general solution is

$$y(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t}, \quad c_1, c_2 \text{ arbitrary constants}$$

$\Delta = 0$: $r_1 = r_2 = r$ real, $y_1(t) = e^{rt}$ and $y_2(t) = te^{rt}$ are 2 linearly independent solutions of (9.8). The general solution is

$$y(t) = c_1 e^{rt} + c_2 te^{rt}, \quad c_1, c_2 \text{ arbitrary constants}$$

$\Delta < 0$: r_1, r_2 complex conjugates, $r_1 = \alpha + i\mu$ and $r_2 = \alpha - i\mu$. $y_1(t) = e^{\alpha t}\cos(\mu t)$ and $y_2(t) = e^{\alpha t}\sin(\mu t)$ are 2 linearly independent solutions of (9.8). The general solution is

$$y(t) = c_1 e^{\alpha t}\cos(\mu t) + c_2 e^{\alpha t}\sin(\mu t), \quad c_1, c_2 \text{ arbitrary constants}$$

Note that an $n^{th}-$order linear homogeneous differential equation always has a solution equal to zero $y(t) \equiv 0$. When the equation has constant coefficients, the stability of the zero solution is stable (when a solution to an IVP will tend to zero). Stability of the zero solution depends on the eigenvalues, the roots of λ_i of the characteristic equation.

Theorem 9.2.10. *If all of the roots of the characteristic polynomial $P(\lambda)$ are negative or have negative real part, then given any solution $y(t)$ of the homogeneous differential equation*

$$\frac{d^n y}{dy^n} + a_1(t)\frac{d^{n-1}y}{dy^{n-1}} + \cdots + a_{n-1}(t)\frac{dy}{dy} + a_n(t)y = 0,$$

there exist positive constant M and b such that

$$|y(t)| \leq Me^{-bt} \quad for \quad t > 0$$

and

$$\lim_{t\to\infty} |x(t)| = 0$$

Methodologies to find a particular solution to the nonhonogeneous differential equation:

- Undetermined coefficients
- Variation of parameters

Undetermined coefficients (to find a particular solution $Y(t)$)

- Make an assumption about the form of the particular solution $Y(t)$ but with the coefficients unspecified
- The particular solution has to satisfy the nonhomogeneous equation $\Leftrightarrow$ Substitute the assumed expression of $Y(t)$ in the nonhomogeneous equation $\Rightarrow$ Coefficients to be determined

Conditions for application: if $g(t)$ takes the form: constant function, exponential, polynomial, sine, cosine, any sum or products of such functions.

$$y'' + p(t)y' + q(t)y = g(t)$$

- Make an assumption about the form of the particular solution $Y(t)$: if $g(t)$ is of the form in the left column or is the sum or product of such function, then check a particular solution $Y(t)$ of the corresponding form as indicated in the right column

$\mathbf{g(t)}$	**Assumed form of** $Y(t)$
$\alpha e^{\beta t}$	$ae^{\beta t}$
$\alpha\cos(\omega t) + \beta\sin(\omega t)$	$a\cos(\omega t) + b\sin(\omega t)$
α	a
$\alpha + \beta t$	$a + bt$
$\alpha + \beta t + \gamma t^2$	$a + bt + ct^2$
$\alpha + \beta t + \gamma t^2 + \cdots + \delta t^m$	$a + bt + ct^2 + \cdots + dt^m$

- If coefficients cannot be determined, then no solution of this form exists $\Rightarrow$ modify the assumption
- If $g(t)$ contains terms that duplicate any solutions of the corresponding homogeneous equation $\Rightarrow$ each such term must be multiplied by t^s (s the smallest natural number that eliminates the duplication).

Variation of parameters: $y'' + p(t)y' + q(t)y = g(t)$

i) Find the **general solution of the homogeneous equation** $\Leftrightarrow$ Find 2 linearly independent solutions y_1 and y_2 of the homogeneous equation

ii) Check for the nonhomogeneous equation a solution of the form

$$Y(t) = u_1(t)y_1(t) + u_2(t)y_2(t)$$

where u_1 and u_2 are two functions of t to be determined, and that satisfy the **second condition**

$$u_1'(t)y_1(t) + u_2'(t)y_2(t) = 0$$

iii) Substitute $Y(t)$ in the nonhomogeneous equation and use the second condition to obtain

$$u_1'(t)y_1'(t) + u_2'(t)y_2'(t) = g(t)$$

iv) Solve the system of 2 linear algebraic equation for u_1' and u_2'

$$\begin{cases} u_1'(t)y_1(t) + u_2'(t)y_2(t) &= 0 \\ u_1'(t)y_1'(t) + u_2'(t)y_2'(t) &= g(t) \end{cases} \text{ or } \begin{bmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{bmatrix} \begin{bmatrix} u_1'(t) \\ u_2'(t) \end{bmatrix} = \begin{bmatrix} 0 \\ g(t) \end{bmatrix}$$

$$\Rightarrow u_1'(t) = \frac{-y_2(t)g(t)}{W(y_1, y_2)(t)}, \qquad u_2'(t) = \frac{y_1(t)g(t)}{W(y_1, y_2)(t)}$$

v) Integrate, evaluate the integral omitting the integration constants

$$\Rightarrow u_1(t) = -\int \frac{y_2(t)g(t)}{W(y_1, y_2)(t)}dt, \qquad u_2(t) = \int \frac{y_1(t)g(t)}{W(y_1, y_2)(t)}dt$$

vi) A **particular solution** of the nonhomogeneous equation is

$$Y(t) = -y_1(t)\int \frac{y_2(t)g(t)}{W(y_1, y_2)(t)}dt + y_2(t)\int \frac{y_1(t)g(t)}{W(y_1, y_2)(t)}dt$$

vii) The **general solution** of the nonhomogeneous equation is

$$y(t) = c_1y_1(t) + c_2y_2(t) + Y(t), \qquad c_1,\ c_2 \text{ arbitrary constants}$$

Theorem 9.2.11. *If the functions p, q and g are continuous on an open interval I, and if the functions y_1 and y_2 are linearly independent solutions of the homogeneous equation, $y'' + p(t)y' + q(t)y = 0$, corresponding to the nonhomogeneous equation*

$$y'' + p(t)y' + q(t)y = g(t)$$

then a particular solution of the nonhomogeneous equation is

$$Y(t) = -y_1(t)\int_{t_0}^{t} \frac{y_2(s)g(s)}{W(y_1, y_2)(s)}ds + y_2(t)\int_{t_0}^{t} \frac{y_1(s)g(s)}{W(y_1, y_2)(s)}ds$$

where t_0 is any point in I. The general solution is

$$y(t) = c_1y_1(t) + c_2y_2(t) + Y(t).$$

9.3 Systems of linear equations

A $n-$dimensional linear system

$$\frac{dX}{dt} = AX$$

where A is a $n \times n$ constant matrix with real elements. The general solution of this system is

$$X(t) = e^{At}C$$

where e^{At} is a $n \times n$ matrix known as the fundamental matrix and C ia a $n \times 1$ vector. There exist many methods to compute the matrix exponential. To compute the general solution of the $n-$dimensional system, a straightforward method can be used (similar method used for higher-order constant coefficient DE).

Here the case of a 2-dimensional linear system is studied. Consider a system of first-order linear equations

$$\frac{dX}{dt} = AX \tag{9.9}$$

where $X = (x_1, x_2)^T$ and A is a constant 2×2 matrix. Note that the zero solution $X = 0$ is solution of the differential equation.

Let $X = e^{\lambda t}V$ be a solution of (9.9), then

$$AV = \lambda V$$

where λ is the eigenvalue of A and V is the eigenvector corresponding to λ. The eigenvalues are solutions of

$$\det(A - \lambda I) = 0.$$

Then

$$\lambda^2 - \mathrm{tr}(A)\lambda + \det(A) = 0.$$

Solutions of (9.9) can take 3 different forms

- if eigenvalues are real and distinct

  $$X(t) = c_1 V_1 e^{\lambda_1 t} + c_2 V_2 e^{\lambda_2 t}$$

  with c_1 and c_2 constant.

- if eigenvalues are real and equal

  $$X(t) = c_1 V_1 e^{\lambda_1 t} + c_2 \left(V_1 t e^{\lambda_1 t} + P e^{\lambda_1 t}\right)$$

  with c_1 and c_2 constant. P can be obtained by solving $(A - \lambda_1 I)P = V_1$.

- if eigenvalues are complex conjugate $\lambda_{1,2} = a \pm ib$

  $$X(t) = P_1 e^{at}\cos(bt) + P_2 e^{at}\sin(bt).$$

Behaviors of solutions:

- The origin is asymptotically stable if the eigenvalues of A are negative or have negative real part.
- The origin is stable if the eigenvalues of A are nonpositive or have nonpositive real part.
- The origin is unstable if the eigenvalues of A are positive or have positive real part.

Theorem 9.3.1. *Suppose* $\frac{dX}{dt} = AX$ *where* A *is a constant* 2×2 *matrix with* $\det(A) \neq 0$.
The orign is asymptotically stable iff

$$tr(A) < 0 \quad and \quad \det(A) > 0.$$

The orign is stable iff

$$tr(A) \leq 0 \quad and \quad \det(A) > 0.$$

The origin is unstable iff

$$tr(A) > 0 \quad or \quad \det(A) < 0.$$

In the case of real eigenvalues λ_1 and λ_2, the eigenvectors V_1 and V_2 are directions along which solutions travel toward or away the origin:

- if λ_1 is positive, solutions will travel along V_1 away from the origin.
- if λ_2 is negative, solutions will travel along V_2 toward the origin.

Hence solutions travel in a direction which corresponds to a linear combination of V_1 and V_2

Theorem 9.3.2. *Consider a system of first order linear equations*

$$\frac{dX}{dt} = AX$$

where $X = (x_1, x_2)^T$ *and* A *is a* $2 \times 2-$*matrix.*

- *if* $\det A > 0$ *and* $trA^2 - 4 \det A \geq 0$ *then the origin is a node (real eigenvalues having the same signs); a stable node if* $trA < 0$ *(real* $\lambda_{1,2} < 0$*), and an unstable node if* $trA > 0$ *(real* $\lambda_{1,2} > 0$*).*
- *if* $\det A < 0$ *then the origin is a saddle (real eigenvalues have opposite signs,* $\lambda_1 \lambda_2 < 0$*).*
- *if* $\det A > 0$ *and* $trA^2 - 4 \det A < 0$ *and* $trA \neq 0$*, the origin is a spiral (complex conjugate with nonzero real part); it is stable if* $trA < 0$ *(negative real part) and unstable if* $trA > 0$ *(positive real part).*
- *if* $\det A > 0$ *and* $trA = 0$ *then the origin is a center (purely imaginary eigenvalues* $\lambda_{1,2} = \pm ib$*).*

9.4 Linear systems of ODE

Definition 9.4.1 (Linear ODE)**.** *A* linear *ODE is a differential equation taking the form*

$$\frac{d}{dt}x = A(t)x + B(t), \tag{9.10}$$

where $A(t) \in \mathcal{M}_n(\mathbb{R})$ *with continuous entries,* $B(t) \in \mathbb{R}^n$ *with real valued, continuous coefficients, and* $x \in \mathbb{R}^n$*. The associated IVP takes the form*

$$\begin{aligned} \frac{d}{dt}x &= A(t)x + B(t) \\ x(t_0) &= x_0. \end{aligned} \tag{9.11}$$

Types of systems:

- $x' = A(t)x + B(t)$ is linear nonautonomous ($A(t)$ depends on t) nonhomogeneous (also called *affine* system).
- $x' = A(t)x$ is linear nonautonomous homogeneous.
- $x' = Ax + B$, that is, $A(t) \equiv A$ and $B(t) \equiv B$, is linear autonomous nonhomogeneous (or affine autonomous).
- $x' = Ax$ is linear autonomous homogeneous.
- If $A(t+T) = A(t)$ for some $T > 0$ and all t, then linear periodic.

Theorem 9.4.2 (Existence and Uniqueness)**.** *Solutions to* (9.11) *exist and are unique on the whole interval over which A and B are continuous. In particular, if A, B are constant, then solutions exist on $\mathbb{R}$.*

Autonomous linear systems Consider the autonomous affine system

$$\frac{d}{dt}x = Ax + B, \tag{9.12}$$

and the associated homogeneous autonomous system

$$\frac{d}{dt}x = Ax. \tag{9.13}$$

9.4.1 Exponential of a matrix

Definition 9.4.3 (Matrix exponential)**.** *Let $A \in \mathcal{M}_n(\mathbb{K})$ with $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. The* exponential *of A, denoted e^{At}, is a matrix in $\mathcal{M}_n(\mathbb{K})$, defined by*

$$e^{At} = \mathbb{I} + \sum_{k=1}^{\infty} \frac{t^k}{k!} A^k,$$

where $\mathbb{I}$ is the identity matrix in $\mathcal{M}_n(\mathbb{K})$.

Properties of the matrix exponential

- $e^{At_1}e^{At_2} = e^{A(t_1+t_2)}$ for all $t_1, t_2 \in \mathbb{R}$.
- $Ae^{At} = e^{At}A$ for all $t \in \mathbb{R}$. [A and e^{At} commute]
- $(e^{At})^{-1} = e^{-At}$ for all $t \in \mathbb{R}$.

Theorem 9.4.4. *The unique solution ϕ of* (9.13) *with initial condition $\phi(t_0) = x_0$ is given by*

$$\phi(t) = e^{A(t-t_0)}x_0.$$

9.4.2 Computing the matrix exponential

Let P be a nonsingular matrix in $\mathcal{M}_n(\mathbb{R})$. We transform the IVP

$$\begin{aligned} \frac{d}{dt}x &= Ax \\ x(t_0) &= x_0, \end{aligned} \tag{9.14}$$

using the transformation $x = Py$ or $y = P^{-1}x$. The dynamics of y is

$$\begin{aligned} y' &= (P^{-1}x)' \\ &= P^{-1}x' \\ &= P^{-1}Ax \\ &= P^{-1}APy. \end{aligned}$$

The initial condition is $y_0 = P^{-1}x_0$. We have thus transformed IVP (9.14) into

$$\begin{aligned} \frac{d}{dt}y &= P^{-1}APy \\ y(t_0) &= P^{-1}x_0. \end{aligned} \tag{9.15}$$

By Theorem 9.4.4, the solution of (9.15) is given by

$$\psi(t) = e^{P^{-1}AP(t-t_0)}P^{-1}x_0,$$

and since $x = Py$, the solution to (9.14) is given by

$$\phi(t) = Pe^{P^{-1}AP(t-t_0)}P^{-1}x_0.$$

So everything depends on $P^{-1}AP$.

9.4.3 Matrix exponential – Diagonalizable case

Before we begin, recall the following definitions and results (see for example [5]).

Definition 9.4.5. *Let $A, B \in \mathcal{M}_n(\mathbb{K})$. A and B are* similar *if there exists $P \in \mathcal{M}_n(\mathbb{K})$ such that $P^{-1}AP = B$.*

Theorem 9.4.6. *A matrix $A \in \mathcal{M}_n(\mathbb{K})$ is similar to a diagonal matrix if and only if A has n linearly independent eigenvectors. In particular, if A has distinct eigenvalues, then A is similar to a diagonal matrix.*

Assume P nonsingular in $\mathcal{M}_n(\mathbb{R})$ such that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

with all eigenvalues $\lambda_1, \ldots, \lambda_n$ different. We have

$$e^{P^{-1}AP} = \mathbb{I} + \sum_{k=1}^{\infty} \frac{t^k}{k!} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}^k$$

For a (block) diagonal matrix M of the form

$$M = \begin{pmatrix} m_{11} & & 0 \\ & \ddots & \\ 0 & & m_{nn} \end{pmatrix}$$

there holds

$$M^k = \begin{pmatrix} m_{11}^k & & 0 \\ & \ddots & \\ 0 & & m_{nn}^k \end{pmatrix}$$

Therefore,

$$\begin{aligned} e^{P^{-1}AP} &= \mathbb{I} + \sum_{k=1}^{\infty} \frac{t^k}{k!} \begin{pmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_n^k \end{pmatrix} \\ &= \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} \end{aligned}$$

And so the solution to (9.14) is given, in the case that A is diagonalizable, by

$$\phi(t) = P \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} P^{-1} x_0. \tag{9.16}$$

9.4.4 Matrix exponential – Nondiagonalizable case

Definition 9.4.7 (Generalized eigenvectors)**.** *Let $A \in \mathcal{M}_n(\mathbb{R})$. Suppose λ is an eigenvalue of A with multiplicity $m \leq n$. Then, for $k = 1, \ldots, m$, any nonzero solution v of*

$$(A - \lambda \mathbb{I})^k v = 0$$

is called a generalized eigenvector *of A.*

Definition 9.4.8 (Nilpotent matrix)**.** *Let $A \in \mathcal{M}_n(\mathbb{R})$. A is* nilpotent *(of order k) if $A^j \neq 0$ for $j = 1, \ldots, k-1$, and $A^k = 0$.*

Theorem 9.4.9 (Jordan normal form)**.** *Let $A \in \mathcal{M}_n(\mathbb{R})$ have eigenvalues $\lambda_1, \ldots, \lambda_n$, repeated according to their multiplicities.*

- *Then there exists a basis of generalized eigenvectors for $\mathbb{R}^n$.*

- *And if $\{v_1, \dots, v_n\}$ is any basis of generalized eigenvectors for $\mathbb{R}^n$, then the matrix $P = [v_1 \cdots v_n]$ is invertible, and A can be written as*

$$A = S + N,$$

where

$$P^{-1}SP = diag(\lambda_j),$$

the matrix $N = A - S$ is nilpotent of order $k \leq n$, and S and N commute, i.e., $SN = NS$.

Another formulation of the same theorem, which is "self-contained", is as follows.

Theorem 9.4.10. *Let $A \in \mathcal{M}_n(\mathbb{R})$ with eigenvalues $\lambda_1, \dots, \lambda_s$ with multiplicities $m_1, \dots, m_s$:*

$$\det(A - \lambda I) = \prod_{j=1}^{s} (\lambda - \lambda_j)^{m_j}.$$

Then A is similar to a matrix of the form

$$J = diag(\Lambda_1, \dots, \Lambda_s),$$

where each block λ_i is an $m_i \times m_i$-matrix of the form

$$\Lambda_i = \begin{pmatrix} \lambda_i & * & 0 & \cdots & 0 \\ 0 & \lambda_i & * & \cdots & 0 \\ & & & & \\ 0 & & & \lambda_i & * \\ & & & 0 & \lambda_i \end{pmatrix}.$$

The Jordan canonical form is

$$P^{-1}AP = \begin{pmatrix} J_0 & & 0 \\ & \ddots & \\ 0 & & J_s \end{pmatrix},$$

so we use the same property as before (but with block matrices now), and

$$e^{P^{-1}APt} = \begin{pmatrix} e^{J_0 t} & & 0 \\ & \ddots & \\ 0 & & e^{J_s t} \end{pmatrix}.$$

The first block in the Jordan canonical form takes the form

$$J_0 = \begin{pmatrix} \lambda_0 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}$$

and thus, as before,

$$e^{J_0 t} = \begin{pmatrix} e^{\lambda_0 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_k t} \end{pmatrix}.$$

Other blocks J_i are written as

$$J_i = \lambda_{k+i}\mathbb{I} + N_i,$$

with $\mathbb{I}$ the $n_i \times n_i$ identity and N_i the $n_i \times n_i$ nilpotent matrix

$$N_i = \begin{pmatrix} 0 & 1 & 0 & 0 \\ & & \ddots & \\ & & & 1 \\ 0 & & & 0 \end{pmatrix}.$$

$\lambda_{k+i}\mathbb{I}$ and N_i commute, and thus

$$e^{J_i t} = e^{\lambda_{k+i} t} e^{N_i t}.$$

Since N_i is nilpotent, $N_i^k = 0$ for all $k \geq n_i$. Therefore, the series $e^{N_i t}$ terminates and

$$e^{J_i t} = e^{\lambda_{k+i} t} \begin{pmatrix} 1 & t & \cdots & \frac{t^{n_i-1}}{(n_i-1)!} \\ 0 & 1 & \cdots & \frac{t^{n_i-2}}{(n_i-2)!} \\ & & & \\ 0 & & & 1 \end{pmatrix}.$$

Theorem 9.4.11. *For all $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$, there is a unique solution $x(t)$ to (9.14) defined for all $t \in \mathbb{R}$. Each coordinate function of $x(t)$ is a linear combination of functions of the form*

$$t^k e^{\alpha t}\cos(\beta t) \quad \text{and} \quad t^k e^{\alpha t}\sin(\beta t)$$

where $\alpha + i\beta$ is an eigenvalue of A and k is less than the algebraic multiplicity of the eigenvalue.

Theorem 9.4.12. *Under conditions of the Jordan normal form Theorem 9.4.9, the linear system $x' = Ax$ with initial condition $x(0) = x_0$, has solution*

$$x(t) = P\,diag\left(e^{\lambda_j t}\right) P^{-1} \left(\mathbb{I} + Nt + \cdots \frac{t^k}{k!} N^k\right) x_0.$$

The result is particularly easy to apply in the following case, where we do not need the matrix P (the basis of generalized eigenvectors).

Theorem 9.4.13 (Case of an eigenvalue of multiplicity n)**.** *Suppose that λ is an eigenvalue of multiplicity n of $A \in \mathcal{M}_n(\mathbb{R})$. Then $S = diag(\lambda)$, and the solution of $x' = Ax$ with initial value x_0 is given by*

$$x(t) = e^{\lambda t} \left(\mathbb{I} + Nt + \cdots \frac{t^k}{k!} N^k\right) x_0.$$

Finally, we give the following variation of constants formula, using the matrix exponential.

Theorem 9.4.14 (Variation of constants formula)**.** *Consider the IVP*

$$x' = Ax + B(t) \tag{9.17a}$$

$$x(t_0) = x_0, \tag{9.17b}$$

where $B : \mathbb{R} \to \mathbb{R}^n$ a smooth function on $\mathbb{R}$, and let $e^{A(t-t_0)}$ be matrix exponential associated to the homogeneous system $x' = Ax$. Then the solution ϕ of (9.17) is given by

$$\phi(t) = e^{A(t-t_0)} x_0 + \int_{t_0}^{t} e^{A(t-s)} B(s) ds. \tag{9.18}$$

9.5 The Laplace transform

Definition 9.5.1 (Laplace transform)**.** *Let $f(t)$ be a function defined for $t \geq 0$. The* Laplace transform *of f is the function $F(s)$ defined by*

$$F(s) = \mathcal{L}\{f(t)\} = \int_0^\infty e^{-st} f(t) dt.$$

The Laplace transform is a linear operator:

$$\mathcal{L}\{af(t) + bg(t)\} = a\mathcal{L}\{f(t)\} + b\mathcal{L}\{g(t)\}.$$

We have the following rules of transformation:

t-domain	s-domain
$af(t) + bg(t)$	$aF(s) + bG(s)$
$tf(t)$	$-F'(s)$
$t^n f(t)$	$(-1)^n F^{(n)}(s)$
f'	$sF(s) - f(0)$
f''	$s^2F(s) - sf(0) - f'(0)$
$f^{(n)}$	$s^nF(s) - s^{n-1}f(0) - \cdots - f^{(n-1)}(0)$
$\frac{f(t)}{t}$	$\int_s^\infty F(u)du$
$\int_0^t f(u)du = u(t) * f(t)$	$\frac{1}{s}F(s)$
$f(at)$	$\frac{1}{\lvert a\rvert}F\left(\frac{s}{a}\right)$
$e^{at}f(t)$	$F(s-a)$
$f(t-a)u(t-a)$	$e^{-as}F(s)$
$f(t) * g(t)$	$F(s)G(s)$

Here $f^{(n)}$ represents the nth derivative, not the nth iterate, and $*$ is the convolution product,

$$(f * g)(t) = \int_a^b f(s)g(t-s)ds.$$

Dirac delta – Heaviside function In the table on the following slide,

- $\delta(t)$ is the Dirac delta,
$$\delta(t) = \begin{cases} \infty & \text{if } t = 0 \\ 0 & \text{if } t \neq 0. \end{cases}$$
- $H(t)$ is the Heaviside function,
$$H(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t > 0. \end{cases}$$

Note that $H(t) = \int_{-\infty}^t \delta(s)ds$.

Transforms of common functions

t-domain	s-domain
$\delta(t)$	1
$\delta(t-\tau)$	$e^{-\tau s}$
$H(t)$	$\frac{1}{s}$
$H(t-\tau)$	$\frac{e^{-\tau s}}{s}$
$\frac{t^n}{n!}H(t)$	$\frac{1}{s^{n+1}}$
$e^{-\alpha t}H(t)$	$\frac{1}{s+\alpha}$
$\sin(\omega t)H(t)$	$\frac{\omega}{s^2+\omega^2}$
$\cos(\omega t)H(t)$	$\frac{s}{s^2+\omega^2}$

Definition 9.5.2. *Given a function $F(s)$, if there exists $f(t)$, continuous on $[0,\infty)$ and such that*

$$\mathcal{L}\{f\} = F,$$

then $f(t)$ is the inverse Laplace transform *of $F(s)$, and is denoted $f = \mathcal{L}^{-1}\{F\}$.*

Theorem 9.5.3. *The inverse Laplace transform is a linear operator. Assume that $\mathcal{L}^{-1}\{F_1\}$ and $\mathcal{L}^{-1}\{F_2\}$ exist, then*

$$\mathcal{L}^{-1}\{aF_1 + bF_2\} = a\mathcal{L}^{-1}\{F_1\} + b\mathcal{L}^{-1}\{F_2\}.$$

Solving differential equations using the Laplace transform

i) Take the Laplace transform of both sides of the equation.

ii) Using the initial conditions, deduce an algebraic system of equations in s-space.

iii) Solve the algebraic system in s-space.

iv) Take the inverse Laplace transform of the solution in s-space, to obtain the solution of the differential equation in t-space.

9.6 Systems of nonlinear equations

Theorem 9.6.1. *(Existence and Uniqueness) Assume that F and $\frac{\partial F}{\partial x_i}$ for $i = 1,\dots,n$ are continuous functions of $(x_1, x_2, \dots, x_n)$ on $\mathbb{R}^n$. Then a unique solution exits to the initial value problem*

$$\frac{dX}{dt} = F(X), \quad X(t_0) = X_0$$

for any initial value $X_0 \in \mathbb{R}^n$.

Theorem 9.6.2. *(Hartman-Grobman) Assume that $(\bar{x}, \bar{y})$ is a hyperbolic (all eigenvalues of the Jacobian matrix evaluated at $(\bar{x}, \bar{y})$ have nonzero real part) equilibrium. Then, in a small neighborhood of $(\bar{x}, \bar{y})$, the nonlinear system behaves in a similar manner as the linearized system.*

Theorem 9.6.3. *Assume the first-order partial derivatives of f and g are continuous in some open set containing the equilibrium $(\bar{x}, \bar{y})$ of the system*

$$\begin{aligned} \frac{dx}{dt} &= f(x,y), \\ \frac{dy}{dt} &= g(x,y). \end{aligned}$$

Then the equilibrium is locally asymptotically stable if

$$trJ < 0 \quad and \quad \det J > 0,$$

where J is the Jacobian matrix evaluated at the equilibrium $(\bar{x}, \bar{y})$. In addition, the equilibrium is unstable if either $trJ > 0$ or $\det J < 0$.

As the linearization is only an approximation of the nonlinear system, the nonlinear system may behave differently from the linear system in 3 cases:

- $\det(J) = 0$: there exists at least one zero eigenvalue, then the equilibrium in the nonlinear system may be a node, a saddle or a spiral.
- $\mathrm{tr}(J) = 0$ and $\det(J) > 0$: eigenvalues are purely imaginary. The equilibrium in the nonlinear system may be a spiral or a center.
- $\mathrm{tr}(J)^2 = 4\det(J)$: in the nonlinear system the equilibrium may be a node or a spiral.

Theorem for $n-$dimensional system:

Theorem 9.6.4. *Suppose $dX/dt = F(X)$ is a nonlinear first-order autonomous system with an equilibrium $\bar{X}$. Denote the Jacobian matrix of F evaluated at $\bar{X}$ as $J(\bar{X})$. If the characteristic equation of the Jacobian matrix $J(\bar{X})$,*

$$\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n = 0$$

satisfies the conditions of the Routh-Hurwitz criteria, that is, the determinants of all the Hurwitz matrices are positive $\det(H_j) > 0$, $j = 1, \ldots, n$ then the equilibrium $\bar{X}$ is L.A.S. If $\det(H_j) < 0$ for some $j = 1, \ldots, n$ then the equilibrium $\bar{X}$ is unstable.

9.7 Phase plane analysis

To study the qualitative behavior of a system without solving it.

$$\begin{aligned} \frac{dx}{dt} &= f(x,y), \\ \frac{dy}{dt} &= g(x,y) \end{aligned}$$

Solutions curves (trajectories) $(x(t), y(t))$ are parametric equations with t as an parameter.

At any point (x, y),

$$\frac{dy}{dx} = \frac{g(x,y)}{f(x,y)}$$

is the slope of the trajectory in the $xy-$plane and the tangent vector that gives the direction of the trajectory is $(f(x,y), g(x,y))$. The collection of vectors evaluated at any point of the $xy-$plane defines the direction field.

Definition 9.7.1. *The $x-$nullcline for*

$$\begin{aligned} \frac{dx}{dt} &= f(x,y), \\ \frac{dy}{dt} &= g(x,y) \end{aligned}$$

is the set of all points in the $xy-$plane satisfying $f(x,y)=0$.

The $y-$nullcline for

$$\begin{aligned} \frac{dx}{dt} &= f(x,y), \\ \frac{dy}{dt} &= g(x,y) \end{aligned}$$

is the set of all points in the $xy-$plane satisfying $g(x,y)=0$.

At any intersection of $x-$nullcline and $y-$nullcline, there is an equilibrium point.

On the $x-$nullcline, all vectors are vertical. On the $y-$nullcline, all vectors are horizontal. We need to check if the direction of flow is up or down on the $x-$nullcline, and if the direction of flow is left or right on the $y-$nullcline.

9.8 Bifurcations

Mathematical models have many parameters. When parameter values change, a change in the behavior of the solution can be expected. If the variation of a parameter change the qualitative behavior of the solution, there is a bifurcation.

Consider the differential equation

$$\frac{dx}{dt} = f(x,\mu), \quad x \in \mathbb{R} \quad \mu \in \mathbb{R}.$$

where μ is the parameter.

Definition 9.8.1. *$\bar{x}$ is a bifurcation point and $\bar{\mu}$ is a bifurcation value if*

$$f(\bar{x},\bar{\mu}) = 0, \quad \textit{and} \quad \frac{\partial}{\partial x} f(\bar{x},\bar{\mu}) = 0.$$

Different types of bifurcations in the case of scalar differential equations:

- Saddle-node bifurcation
- Transcritical bifurcation
- Pitchfork bifurcation

These types of bifurcations also occur in higher-dimensional system.

A fourth type of bifurcation can occurs in systems consisting of two or more equations: the Hopf bifurcation.

Consider a system of autonomous DEs:

$$\begin{aligned} \frac{dx}{dt} &= f(x,y,r) \\ \frac{dy}{dt} &= g(x,y,r) \end{aligned}$$

f and g depends on the parameter r. Assume that $(\bar{x}(r), \bar{y}(r))$ is an equilibrium and the Jacobian evaluated at this equilibrium has eigenvalues $\alpha(r) \pm i\beta(r)$.

A change of stability occurs at $r = \bar{r}$ where $\alpha(\bar{r}) = 0$. If $\alpha(r) < 0$ for $r < \bar{r}$ and $\alpha(r) > 0$ for $r > \bar{r}$ (with $\beta(r) \neq 0$), then the equilibrium changes from a stable spiral to an unstable spiral when r passes through $\bar{r}$. The Hopf Theorem states that there is a periodic orbit near $r = \bar{r}$ from any neighborhood of the equilibrium is $\mathbb{R}^2$. Then r is the parameter of bifurcation and the bifurcation value is $\bar{r}$.

Theorem 9.8.2. *(Hopf bifurcation) Consider the system*

$$\begin{aligned}\frac{dx}{dt} &= f(x, y, r)\\ \frac{dy}{dt} &= g(x, y, r)\end{aligned}$$

where $f(x, y, r)$ and $g(x, y, r)$ are continous and differentiable. The system has an equilibrium $(\bar{x}(r), \bar{y}(r))$, and the Jacobian of the system evaluated at this parameter-dependent equilibrium is $J(r)$. The Jacobian matrix $J(r)$ has eigenvalues $\alpha(r) \pm i\beta(r)$.

Assume that there exists a value $\bar{r}$ called the bifurcation value, such that $\alpha(\bar{r}) = 0$ and $\beta(\bar{r}) \neq 0$, and as r is varied through, $\bar{r}$, the real part of the eigenvalues change signs

$$\frac{d\alpha}{dr}_{r=\bar{r}} \neq 0.$$

Given these following hypotheses, the following possibilities arise:

- *At $r = \bar{r}$, a center is created at the equilibrium, and thus infinitely many neutrally stable concentric closed orbits surround $(\bar{x}(r), \bar{y}(r))$.*
- *There is a range of r values that $\bar{r} < r < c$ for which a single closed orbit (a limit cycle) surround $(\bar{x}(r), \bar{y}(r))$. As r is varied the diameter of the limit cycle changes in proportion to $\sqrt{|r - \bar{r}|}$. There are no other closed orbits near $(\bar{x}(r), \bar{y}(r))$. Since the limit cycle exists for values above $\bar{r}$, this phenomenon is called a supercritical bifurcation.*
- *There is a range of values such that $d < r < \bar{r}$ for which a single closed orbit (a limit cycle) surround $(\bar{x}(r), \bar{y}(r))$. Since the limit cycle exists for values below $\bar{r}$, this phenomenon is called a supercritical bifurcation.*

Chapter 10

A few epidemic models

Introduction to the analysis of nonlinear systems of ordinary differential equations

Parameters:

- β transmission rate
- b rate of a birth
- d rate of death
- γ rate of recovery; $1/\gamma$ is the average length of the infectious period when there are no death.
- $1/(\gamma + b)$ is the average length of the infectious period when deaths are included.
- ν rate of loss of immunity; $1/\nu$ average length of immunity.

Definition 10.0.3. *The incidence is the rate at which infections occur. The incidence function is defined as $f(I, S) = \lambda(I)S$ where $\lambda(I)$ is the force of infection (probability of a given susceptible contracts the disease).*

Some incidence functions

- Mass action: infectives and susceptible mixed completely with each other, $f(I, S) = \beta IS$.
- Proportional incidence (pseudo mass action): $f(I, S) = \beta \frac{I}{N} S$
- Refuge effect:
$$f(I, S) = \begin{cases} \beta I(N - I/q) & I < qN \\ 0 & I \geq qN \end{cases}$$
where $0 < q < 1$ is the proportion of population potentially suceptible because of spatial or other heterogeneities.

- For vector-borne disease: Criss-cross infection (the vector infecting the host and the host then infecting another vector)
- ...

Definition 10.0.4. *The prevalence of a disease in a population is the fraction infected.*

Definition 10.0.5. *The basic reproduction number $\mathcal{R}_0$ is the average number of secondary infections caused by one infectious individual in a totally susceptible population during the individual's infectious period.*

Figure 10.1: Reproduction number $\mathcal{R}_0 = 2$.

Magnitude of the basic reproductive number gives an indication of the difficulty in controlling an epidemic or eradicating the disease: the larger the value of $\mathcal{R}_0$, the harder it is to control.

10.1 SIS model without vital dynamics

Consider a disease that confers no immunity. In this case, individuals are either

- *susceptible* to the disease, with the number of such individuals at time t denoted by $S(t)$,
- or *infected* by the disease (and are also *infective* in the sense that they propagate the disease), with the number of such individuals at time t denoted by $I(t)$.

We want to model the evolution with time of S and I (t is omitted unless necessary).

Hypotheses

- Individuals recover from the disease at the *per capita* rate γ.
- The disease does not confer immunity.
- There is no birth or death.
- Infection is of *standard incidence* type, $\beta = SI/N$.

(for details, see Chapter on *residence time*)

The evolution of $I(t)$ is described by the following equation (see slides on *residence time*):

$$I' = \beta \frac{(N-I)I}{N} - \gamma I.$$

Develop and reorder the terms, giving

$$I' = (\beta - \gamma)I - \frac{\beta}{N} I^2 \tag{10.1}$$

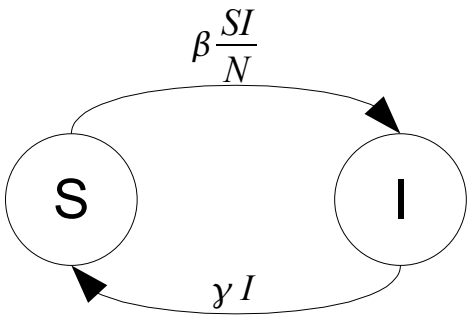


Figure 10.2: Flow diagram of the model

This is a logistic-type equation. It can be solved as a Bernoulli equation or as a separable equation, giving, for an initial number of infectives $I(0) = I_0$,

$$I(t) = \frac{(\beta - \gamma)NI_0}{(\beta - \gamma)Ne^{-(\beta-\gamma)t} + \beta I_0 \left(1 - e^{-(\beta-\gamma)t}\right)}$$

From $S = N - I$, we deduce that the solution $(S(t), I(t))$ for the complete system, with initial condition $S(0) + I(0) = S_0 + I_0 = N$ is, for $t \geq 0$,

$$S(t) = N - \frac{(\beta - \gamma)NI_0}{(\beta - \gamma)Ne^{-(\beta-\gamma)t} + \beta I_0 \left(1 - e^{-(\beta-\gamma)t}\right)}$$

and

$$I(t) = \frac{(\beta - \gamma)NI_0}{(\beta - \gamma)Ne^{-(\beta-\gamma)t} + \beta I_0 \left(1 - e^{-(\beta-\gamma)t}\right)}$$

10.1.1 Behavior of the solutions

Consider only I for the moment.

$$I(t) = \frac{(\beta - \gamma)NI_0}{(\beta - \gamma)Ne^{-(\beta-\gamma)t} + \beta I_0 \left(1 - e^{-(\beta-\gamma)t}\right)}$$

So

- If $\beta - \gamma > 0$, then $e^{-(\beta-\gamma)t} \to 0$ as $t \to \infty$, and therefore

$$\lim_{t\to\infty} I(t) = \frac{(\beta - \gamma)NI_0}{\beta I_0} = \frac{\beta - \gamma}{\beta} N = \left(1 - \frac{\gamma}{\beta}\right) N.$$

- If $\beta - \gamma < 0$, then $e^{-(\beta-\gamma)t} \to \infty$ at $t \to \infty$. This implies that the denominator in $I(t)$ tends to $-\infty$ as $t \to \infty$, and so

$$\lim_{t\to\infty} I(t) = 0, \text{ with } I(t) > 0 \text{ for all } t.$$

- If $\beta = \gamma$, then $I(t) = 0$ for all t.

10.1.2 The basic reproduction number

Define the *basic reproduction number* (the average number of people that an infectious individual will infect, when introduced in a population of susceptibles) as

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

We have

$$(\mathcal{R}_0 < 1 \Leftrightarrow (\beta - \gamma) < 0) \text{ and } (\mathcal{R}_0 > 1 \Leftrightarrow (\beta - \gamma) > 0)\,.$$

Therefore, previous cases can be rewritten

- If $\mathcal{R}_0 < 1$, then $\lim_{t\to\infty} I(t) = 0$.
- If $\mathcal{R}_0 > 1$, then

$$lim_{t\to\infty} I(t) = \left(1 - \frac{1}{\mathcal{R}_0}\right) N.$$

(The case $\mathcal{R}_0 = 1$ is usually omitted.) To plot this in Maple, use the commands

```
> f:=R->piecewise(R<1,0,R>1,(1-1/R)*1000);
> plot(f(R),R=0..10);
```

This gives the result shown in Figure 10.3.

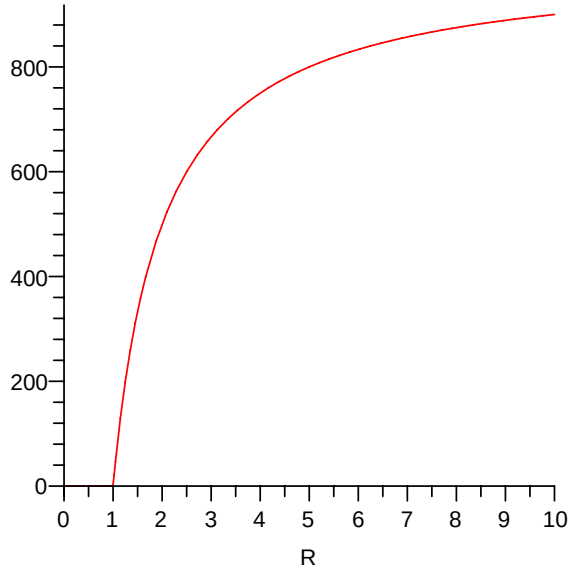


Figure 10.3: Bifurcation diagram showing the number of infectives in the epidemic model, as a function of $\mathcal{R}_0$.

10.2 SIR model of Kermack and McKendrick

In 1927, Kermack and McKendrick started publishing a series of papers on epidemic models [8, 9, 10]. In the first of their papers [8], they have the following model as a particular case:

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \gamma I \\ R' &= \gamma I \end{aligned} \tag{10.2}$$

Analyzing the system First, note (as KMK) that the total population in the system is constant. This is deduced from the fact that

$$N' = (S + I + R)' = -\beta SI + \beta SI - \gamma I + \gamma I = 0.$$

Since this is true for all values of t, we have N constant.

Let us ignore the R equation for now. We can compute

$$\frac{dI}{dS} = \frac{dI}{dt}\frac{dt}{dS} = \frac{I'}{S'} = \frac{\gamma}{\beta S} - 1$$

This gives

$$I(S) = S - \frac{\gamma}{\beta}\ln S + K,$$

which, considering the initial condition (S_0, I_0), is,

$$I(S) = S - \frac{\gamma}{\beta}\ln S + I_0 - (S_0 - \frac{\gamma}{\beta}\ln S_0).$$

This gives a curve in the (S, I) plane.

$$I(S) = S - \frac{\gamma}{\beta}\ln S + I_0 - (S_0 - \frac{\gamma}{\beta}\ln S_0).$$

Typically, assume $S \approx N$ and $I > 0$ small. Let us denote $S_\infty = \lim_{t\to\infty} S(t)$.

We want to find the value of S when $I \to 0$. Then

$$I_0 - \frac{\gamma}{\beta}\ln S_0 = S_\infty - \frac{\gamma}{\beta}\ln S_\infty$$

10.3 SIRS models with demography

10.3.1 The SIRS model

- Like KMK, individuals are S, I or R.
- Infection is βSI (mass action) or $\beta SI/N$ (proportional incidence).
- Different interpretation of the R class: R stands for "recovered", individuals who are immune to the disease following recovery.
- Recovery from the disease (movement from I class to R class) occurs at the per capita rate γ.
  (Time spent in I before recovery is exponentially distributed.)
- Immunity can be lost: after some time, R individuals revert back to S individuals.
- Time spent in R class before loss of immunity is exponentially distributed, with mean $1/\nu$.
- There is birth and death of individuals:

– No vertical transmission of the disease (mother to child) or of immunity, so all birth is into the S class.
Birth occurs at the rate Π.

– Individuals in all classes die of at the per capita rate d, i.e., the average life duration is exponentially distributed with mean $1/d$.

– The disease is lethal: infected individuals are subject to additional mortality at the per capita rate δ.

Note that birth and death can have different interpretations:

- birth and death in the classical sense,
- but also, entering the susceptible population and leaving it.

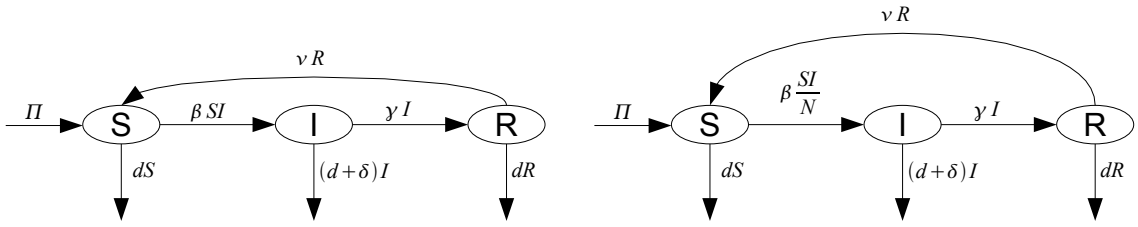


Figure 10.4: The SIRS model with (left) mass action incidence and (right) proportional incidence.

Mass action

$$S' = \Pi + \nu R - \beta SI - dS \tag{10.3a}$$
$$I' = \beta SI - (d + \delta + \gamma)I \tag{10.3b}$$
$$R' = \gamma I - (d + \nu)R \tag{10.3c}$$

Proportional incidence

$$S' = \Pi + \nu R - \beta SI - dS \tag{10.4a}$$
$$I' = \beta SI - (d + \delta + \gamma)I \tag{10.4b}$$
$$R' = \gamma I - (d + \nu)R, \tag{10.4c}$$

where $N = S + I + R$.

We consider (10.3).

10.3.2 Qualitative analysis

Steps of the analysis

i) Assess the well-posedness of the system:

a) Determine whether solutions exist and are unique.

b) Determine whether solutions remain in a realistic region and are bounded.

ii) Study the equilbrium solutions of the system:

a) Find the equilibria of the system.

b) Determine the local stability properties of the equilibria.

c) Determine the global stability properties of the equilibria (**much harder**, often not possible).

iii) In the case of unstable equilibrium points, study the existence of periodic solutions.

Existence and uniqueness of solutions

Theorem 10.3.1 (Cauchy-Lipschitz)**.** *Consider the equation $x' = f(x)$, with $x \in \mathbb{R}^n$, and suppose that $f \in C^1$. Then there exists a unique solution of $x' = f(x)$ such that $x(t_0) = x_0$, where $t_0 \in \mathbb{R}$ and $x_0 \in \mathbb{R}^n$, defined on the largest interval $J \ni t_0$ on which $f \in C^1$.*

Definition 10.3.2 (Equilibrium point)**.** *Consider a differential equation*

$$x' = f(x), \tag{10.5}$$

with $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$. Then x^* is an equilibrium (solution) of (10.5) if $f(x^*) = 0$.

Linearization Consider x^* an equilibrium of (10.5). For simplicity, assume here that $x^* = 0$ (it is always possible to do this, by considering $y = x - x^*$). Taylor's theorem:

$$f(x) = Df(0)x + \frac{1}{2}D^2f(0)(x, x) + \cdots,$$

where $Df(0)$ is the Jacobian matrix of f evaluated at 0.

Stability of equilibria

Definition 10.3.3 (Stable and unstable EP)**.** *Let ϕ_t be the flow of (10.5), assumed to be defined for all $t \in \mathbb{R}$. An equilibrium x^* of (10.5) is (locally)* stable *if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in \mathcal{N}_\delta(x^*)$ and $t \geq 0$, there holds*

$$\phi_t(x) \in \mathcal{N}_\varepsilon(x^*).$$

The equilibrium point is unstable *if it is not stable.*

Definition 10.3.4 (Asymptotically stable EP)**.** *Let ϕ_t be the flow of (10.5) is (locally)* asymptotically stable *if there exists $\delta > 0$ such that for all $x \in \mathcal{N}_\delta(x^*)$ and $t \geq 0$, there holds*

$$\lim_{t\to\infty} \phi_t(x) = x^*.$$

Clearly, Asymtotically Stable $\Rightarrow$ Stable.

Hyperbolic EPs, sinks, sources

Definition 10.3.5 (Sink)**.** *An equilibrium point x^* of* (10.5) *is* hyperbolic *if none of the eigenvalues of the matrix $Df(x^*)$ (Jacobian matrix of f evaluated at x^*) have zero real parts.*

Definition 10.3.6 (Sink)**.** *An equilibrium point x^* of* (10.5) *is a* sink *if all the eigenvalues of the matrix $Df(x^*)$ have negative real parts.*

Definition 10.3.7 (Source)**.** *An equilibrium point x^* of* (10.5) *is a* source *if all the eigenvalues of the matrix $Df(x^*)$ have positive real parts.*

Theorem 10.3.8. *If x^* is a sink of* (10.5) *and for all the eigenvalues λ_j of the matrix $Df(x^*)$*

$$\Re(\lambda_j) < -\alpha < 0,$$

where $\Re(\lambda)$ denotes the real part of λ, then for a given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in \mathcal{N}_\delta(x^*)$, the flow $\phi_t(x)$ of (10.5) *satisfies*

$$\|\phi_t(x) - x^*\| \leq \varepsilon e^{-\alpha t}$$

for all $t \geq 0$.

Theorem 10.3.9. *If x^* is a stable equilibrium point of* (10.5)*, no eigenvalue of $Df(x^*)$ has positive real part.*

Chapter 11

The chemostat

Some notions of phase plane analysis

11.1 The chemostat

A chemostat consists in one main chamber (called a vessel), in which some microorganisms (bacteria, plankton), typically unicellular, are put, together with liquid and nutrient. The contents are stirred, so nutrient and organisms are well-mixed. Organisms consume the nutrient in their environment, which causes them to grow and multiply. Two major modes of operation:

- *Batch* mode: let the whole thing sit.
- *Continuous flow* mode: there is an input of fresh water and nutrient, and an outflow the comprises water, nutrient and organisms, to keep the volume constant.

Chemostats are very popular tools.

- Study of the growth of micro-organisms as a function of nutrient, in a very controlled setting.
- Very good reproducibility of experiments.
- Used in all sorts of settings. Fundamental science, but also, for production of products.

11.2 Batch mode

11.2.1 Model with no cell mortality

We make the following assumptions. Organisms, whose concentration is denoted x, are in the main vessel. Limiting substrate has a concentration in the vessel denoted S. There is homogeneous mixing of the contents of the vessel, so that nutrient is readily available to all organisms at the same concentration. Therefore, spatial aspects can be neglected. Organisms uptake nutrient at the rate $\mu(S)$, a function of the concentration of nutrient around them. First, we assume no death of organisms. The model then is

$$S' = -\mu(S)x \tag{11.1a}$$

$$x' = \mu(S)x \tag{11.1b}$$

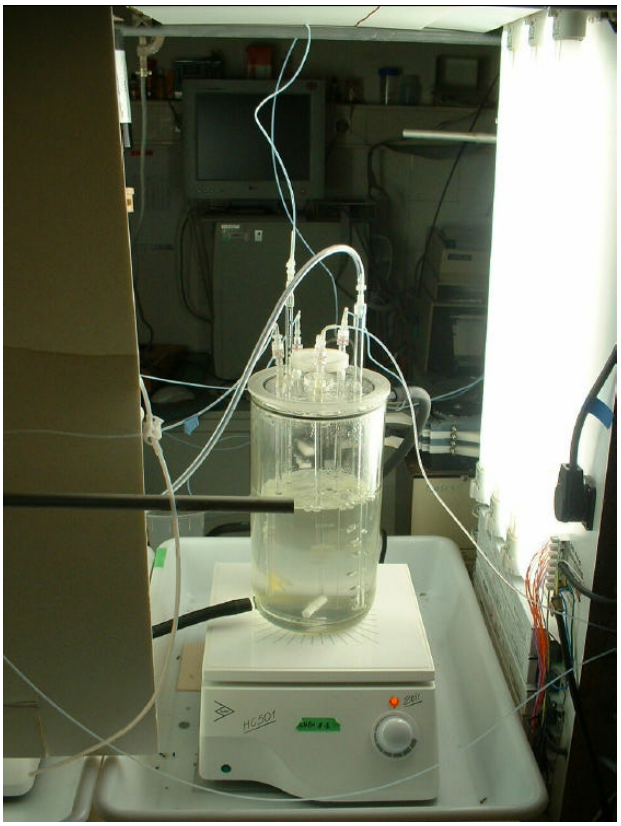


Figure 11.1: A chemostat operating at the Laboratoire Océanographique de Villefranche sur Mer, France.

with initial conditions $S(0) \geq 0$ and $x(0) > 0$, and where $\mu(S)$ is such that

- $\mu(0) = 0$ (no substrate implies no growth)
- $\mu(S) \geq 0$ for all $S \geq 0$
- $\mu(S)$ bounded for $S \geq 0$

A typical form for $\mu(S)$ is the *Monod* curve,

$$\mu(S) = \mu_{max} \frac{S}{K_S + S}. \tag{11.2}$$

The parameter μ_{max} is the *maximal growth rate*, while K_S is the half-saturation constant ($\mu(K_S) = \mu_{max}/2$). See an example in Figure **??**. From now on, we assume a Monod function.

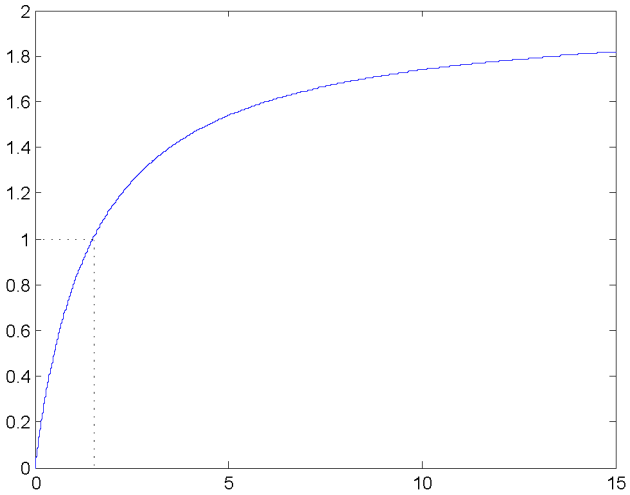


Figure 11.2: A typical Monod growth curve.

In other contexts, this curve is called a Holling Type II function.

11.2.2 Equilibria

To compute the equilibria, suppose $S' = x' = 0$, giving

$$\mu(S)x = -\mu(S)x = 0.$$

This implies $\mu(S) = 0$ or $x = 0$. Note that $\mu(S) = 0 \Leftrightarrow S = 0$, so the system is at equilibrium if $S = 0$ or $x = 0$.

This is a complicated situation, as it implies that there are lines of equilibria ($S = 0$ and any x, and $x = 0$ and any S), so that the equilibria are not *isolated* (arbitrarily small neighborhoods of one equilibrium contain other equilibria), and therefore, studying the linearization is not possible. Here, some analysis is however possible. Consider

$$\frac{dx}{dS} = \frac{dx}{dt}\frac{dt}{dS} = -\frac{\mu(S)x}{\mu(S)x} = -1.$$

This implies that we can find the solution

$$x(S) = C - S,$$

or, supposing the initial condition is $(S(0), x(0)) = (S_0, x_0)$, that is, $x(S_0) = x_0$,

$$x(S) = S_0 + x_0 - S. \tag{11.3}$$

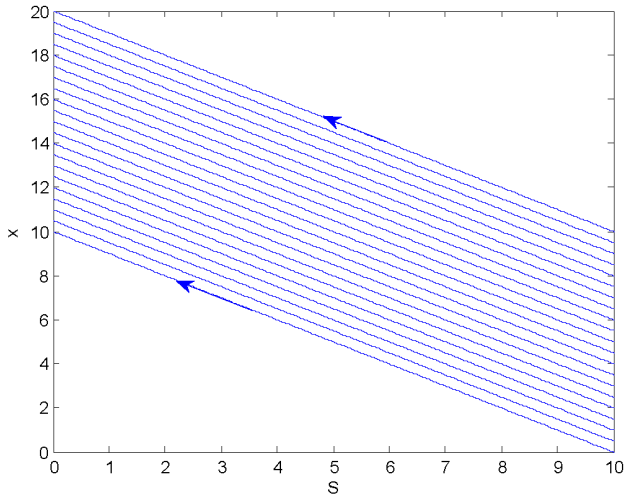


Figure 11.3: Typical solutions obtained from (11.3).

11.2.3 Model with organism death

Assume death of organisms at per capita rate d. Model is

$$S' = -\mu(S)x \tag{11.4a}$$
$$x' = \mu(S)x - dx. \tag{11.4b}$$

We have

$$S' = 0 \Leftrightarrow \mu(S)x = 0$$

and

$$x' = 0 \Leftrightarrow (\mu(S) - d)x = 0.$$

So we have $x = 0$ or $\mu(S) = d$. So $x = 0$ and any value of S, and S such that $\mu(S) = d$ and $x = 0$. One such particular value is $(S, x) = (0, 0)$. This is once again a complicated situation, since there are lines of equilibria. Intuitively, most solutions will go to $(0, 0)$. This is indeed the case (we will not show it).

11.3 Continous flow mode

11.3.1 Modelling principles

General hypotheses are similar to the batch case, except that additionally,

- Limiting substrate (whose concentration in the vessel is denoted S) is input at the rate D and concentration S^0.
- There is an outflow of both nutrient and organisms (at same rate D as input).
- Residence time in device is assumed small compared to lifetime (or time to division) $\Rightarrow$ no death considered.

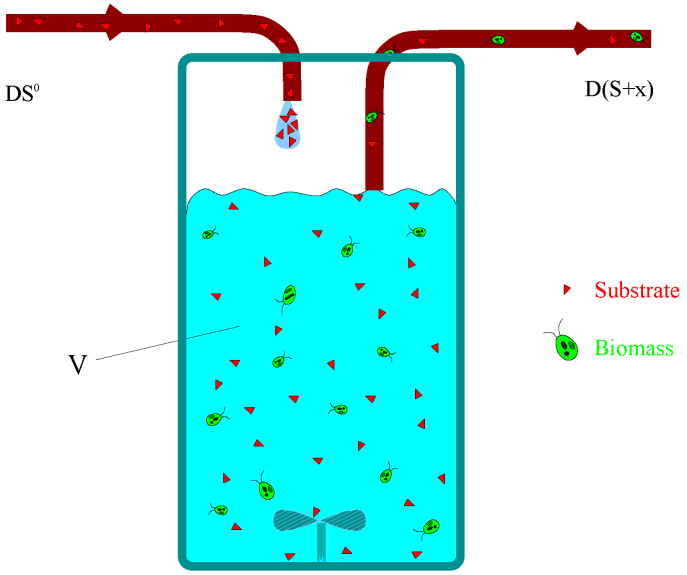


Figure 11.4: A schematic representation of a chemostat operating in continuous flow mode.

11.3.2 Model for continuous flow mode

The model is

$$S' = D(S^0 - S) - \mu(S)x \tag{11.5a}$$
$$x' = \mu(S)x - Dx, \tag{11.5b}$$

with initial conditions $S(0) \geq 0$ and $x(0) \geq 0$, and $D, S^0 > 0$. This is called the *Monod model*.

11.3.3 Finding equilibria

Setting $S' = x' = 0$, we get

$$0 = D(S^0 - S) - \mu_{max}\frac{S}{K_S + S}x$$
$$0 = \left(\mu_{max}\frac{S}{K_S + S} - D\right)x$$

But here, instead of directly computing the values of the equilibrium points, we consider the system in the plane, using nullclines.

11.3.4 Phase plane analysis

Nullclines are the level set 0 of the vector field. If we have

$$x_1' = f_1(x_1, x_2)$$
$$x_2' = f_2(x_1, x_2)$$

then the nullclines for x_1 are the curves defined by

$$\{(x_1, x_2) \in \mathbb{R}^2 : f_1(x_1, x_2) = 0\},$$

those for x_2 are

$$\{(x_1, x_2) \in \mathbb{R}^2 : f_2(x_1, x_2) = 0\}.$$

(Note: In $\mathbb{R}^2$, nullclines are curves.) On the nullcline associated to one state variable, this state variable has zero derivative. Equilibria lie at the intersections of nullclines for both state variables (in $\mathbb{R}^2$).

Nullclines for x Nullclines are given by

$$0 = D(S^0 - S) - \mu_{max}\frac{S}{K_S + S}x \tag{11.6a}$$

$$0 = \left(\mu_{max}\frac{S}{K_S + S} - D\right)x \tag{11.6b}$$

From (11.6b), nullclines for x are $x = 0$ and

$$\mu_{max}\frac{S}{K_S + S} - D = 0$$

Write the latter as

$$\begin{aligned}\mu_{max}\frac{S}{K_S + S} - D = 0 &\Leftrightarrow \mu_{max}S = D(K_S + S)\\ &\Leftrightarrow (\mu_{max} - D)S = DK_S\\ &\Leftrightarrow S = \frac{DK_S}{\mu_{max} - D}\end{aligned}$$

Nullcline for x So, for x, there are two nullclines:

- The line $x = 0$.
- The line $S = \dfrac{DK_S}{\mu_{max} - D}$.

For the line $S = DK_S/(\mu_{max} - D)$, we deduce a condition:

- If $\mu_{max} - D > 0$, that is, if $\mu_{max} > D$, i.e., the maximal growth rate of the cells is larger than the rate at which they leave the chemostat due to washout, then the nullcline intersects the first quadrant.
- If $\mu_{max} < D$, then the nullcline does not intersect the first quadrant.

Nullclines for S Nullclines are given by

$$0 = D(S^0 - S) - \mu_{max}\frac{S}{K_S + S}x \tag{11.6a}$$

$$0 = \left(\mu_{max}\frac{S}{K_S + S} - D\right)x \tag{11.6b}$$

Rewrite (11.6a),

$$\begin{aligned} D(S^0 - S) - \mu_{max}\frac{S}{K_S + S}x = 0 &\Leftrightarrow \mu_{max}Sx = D(S^0 - S)(K_S + S) \\ &\Leftrightarrow x = \frac{D(S^0 - S)(K_S + S)}{\mu_{max}S} \end{aligned}$$

Nullcline for S: S intercept The equation for the nullcline for S is

$$x = \Gamma(S) \triangleq \frac{D}{\mu_{max}}\left(\frac{S^0K}{S} - S + S^0 - K\right)$$

We look for the intercepts. First, S intercept:

$$\begin{aligned} \Gamma(S) = 0 &\Leftrightarrow \frac{S^0K_S}{S} - S + S^0 - K_S = 0 \\ &\Leftrightarrow \frac{S^0K}{S} = S - S^0 + K \\ &\Leftrightarrow S^0K_S = S^2 + (K_S - S^0)S \\ &\Leftrightarrow S^2 + (K - S^0)S - S^0K_S = 0 \end{aligned}$$

Roots of this degree 2 polynomial are $-K_S$ (< 0) and S^0.

Nullcline for S: x intercept x intercept is found at $\Gamma(0)$. But this is not defined (division by $S = 0$), so consider

$$\begin{aligned} \lim_{S\to 0^+}\Gamma(S) &= \lim_{S\to 0^+}\frac{D}{\mu_{max}}\left(\frac{S^0K}{S} - S + S^0 - K\right) \\ &= \frac{D}{\mu_{max}}\left(\lim_{S\to 0^+}\frac{S^0K}{S} - S + S^0 - K\right) \\ &= \frac{D}{\mu_{max}}\left(\lim_{S\to 0^+}\left(\frac{S^0K}{S}\right) + \lim_{S\to 0^+}\left(-S + S^0 - K\right)\right) \\ &= \frac{D}{\mu_{max}}\left(+\infty + S^0 - K\right) \\ &= +\infty. \end{aligned}$$

Maple has a plot function, `implicitplot` (part of the `plots` library), that is very useful for nullclines (d is used instead of D, because maple does not allow to change D without using `unprotect`).

```
> with(plots):
> d := 0.4; S0 := 1; mu := 0.7; K := 2;
> implicitplot(d*(S0-S)-mu*S/(K+S)*x=0,S=0..10,x=0..10)
```

11.3.5 Stability of the equilibria

The computation was done during class.

11.3.6 Conservation of mass

Summing the equations in (11.5), we get

$$(S+x)' = D\left(S^0 - (S+x)\right)$$

Denote $M = S + x$ the total organic mass in the chemostat. Then

$$M' = D(S^0 - M)$$

This is a linear equation in M. Solving it (e.g., integrating factor), we find

$$M(t) = S^0 - e^{-Dt}\left(S^0 - M(0)\right),$$

and so

$$\lim_{t\to\infty} M(t) = S^0.$$

This is called the *mass conservation principle*.

Implication of mass conservation Not as strong as what we had in the SIS epidemic model, where the total number of individuals was constant. Here, the mass is *asymptotically* constant. **But** we can still use it, using the theory of *asymptotically autonomous* differential equations. Too complicated for here, just remember that often, it is *allowed* to use the limit of a variable rather than the variable itself, provided you know that the convergence occurs.

Chapter 12

Traffic flow

Linear cascades
Linear systems
Delay differential equations
Laplace transform

12.1 An ODE model of traffic flow

We want to model a situation with N cars on a straight road with no overtaking, and where a given driver adjusts (instantaneously) their speed on the speed the driver in front of them.

12.1.1 Hypotheses

- N cars in total,
- the road is the x-axis,
- $x_n(t)$ is the position of the nth car at time t,
- $v_n(t) \stackrel{\Delta}{=} x_n'(t)$ is the velocity of the nth car at time t,
- all cars start with the same initial speed v_0 at time $t = 0$.

To make computations easier, we express the velocity of cars in a reference frame moving at the speed u_0. (Remark that here, speed and velocity are equal, since movement is 1-dimensional.) Let

$$u_n(t) = v_n(t) - u_0.$$

Then $u_n(t) = 0$ for $t \leq 0$, and u_n is the speed of the nth car in the moving frame coordinates.

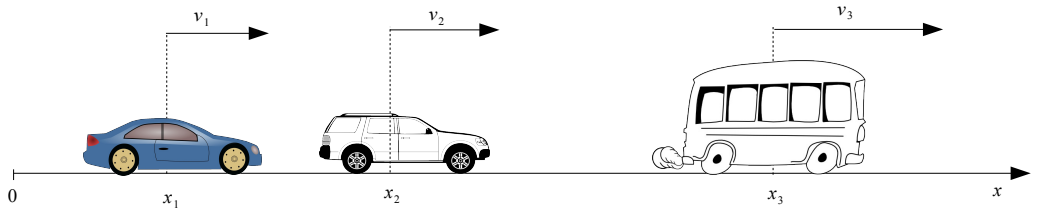


Figure 12.1: The situation described by the model of traffic flow.

12.1.2 Modeling driver behavior

Assume that a driver adjusts their speed according to relative speed between their car and the car in front, and that this adjustment is a linear term, equal to λ for all drivers. This implies the following.

- For the first car, the evolution of speed remains to be determined.
- For the second car,
$$u_2' = \lambda(u_1 - u_2).$$
This means that if the speed of the second car, u_2, is larger than that of the first car, u_1, then the driver of the second car decreases their speed (u_2' has a negative sign). Similarly, if the speed of the second car is smaller than that of the first car, then the term $\lambda(u_2 - u_1)$ is positive, implying that $u_2' > 0$, i.e., the second car speeds up.
- Similarly, the third car adjusts its speeds as a function of the car in front, i.e., car number 2, giving
$$u_3' = \lambda(u_2 - u_3).$$
- Thus, for $n = 1, \ldots, N-1$,
$$u_{n+1}' = \lambda(u_n - u_{n+1}). \tag{12.1}$$

12.1.3 Solving using linear cascades

This can be solved using *linear cascades*: if $u_1(t)$ is known, then

$$u_2' = \lambda(u_1(t) - u_2)$$

is a linear first-order nonhomogeneous equation. The solution (obtained using integrating factors or variation of constants) is

$$u_2(t) = \lambda e^{-\lambda t} \int_0^t u_1(s) e^{\lambda s} ds.$$

Then use this function $u_2(t)$ in u_3' to get $u_3(t)$,

$$u_3(t) = \lambda e^{-\lambda t} \int_0^t u_2(s) e^{\lambda s} ds.$$

Thus

$$\begin{aligned} u_3(t) &= \lambda e^{-\lambda t} \int_0^t u_2(s) e^{\lambda s} ds \\ &= \lambda e^{-\lambda t} \int_0^t \left(\lambda e^{-\lambda s} \int_0^s u_1(q) e^{\lambda q} dq \right) ds \\ &= \lambda^3 e^{-\lambda t} \int_0^t e^{-\lambda s} \int_0^s u_1(q) e^{\lambda q} dq ds. \end{aligned}$$

Continue the process to get the solution.

12.1.4 An example with known first driver behavior

Suppose that the driver of car 1 follows the function

$$u_1(t) = \alpha \sin(\omega t),$$

that is, ω-periodic, 0 at $t = 0$ (we want all cars to start with speed relative to the moving reference frame equal to 0), and with amplitude α. Then

$$\begin{aligned} u_2(t) &= \lambda \alpha e^{-\lambda t} \int_0^t \sin(\omega s) e^{\lambda s} ds \\ &= \lambda \alpha e^{-\lambda t} \left(\frac{\omega - \omega e^{\lambda t} \cos(\omega t) + \lambda e^{\lambda t} \sin(\omega t)}{\lambda^2 + \omega^2} \right) \\ &= \frac{\lambda \alpha}{\lambda^2 + \omega^2} \left(\omega e^{-\lambda t} + \lambda \sin(\omega t) - \omega \cos(\omega t) \right). \end{aligned}$$

When $t \to \infty$, the first term goes to 0 and we are left with a ω-periodic term. Continuing the process,

$$u_3(t) = \frac{\lambda^2 \alpha}{\lambda^2 + \omega^2} e^{-\lambda t} \int_0^t \left(\omega e^{-\lambda s} + \lambda \sin(\omega s) - \omega \cos(\omega s) \right) e^{\lambda s} ds,$$

that is,

$$\begin{aligned} u_3(t) &= \frac{\lambda^2 \alpha}{\lambda^2 + \omega^2} e^{-\lambda t} \left(\omega t + \int_0^t \left(\lambda \sin(\omega s) - \omega \cos(\omega s) \right) e^{\lambda s} ds \right) \\ &= \frac{\lambda^2 \alpha}{\lambda^2 + \omega^2} \left(\omega t + \frac{2\lambda\omega}{\lambda^2 + \omega^2} \right) e^{-\lambda t} \\ &\quad - \frac{\lambda^2 \alpha}{(\lambda^2 + \omega^2)^2} \left(2\lambda\omega \cos(\omega t) - \lambda^2 \sin(\omega t) + \omega^2 \sin(\omega t) \right). \end{aligned}$$

Once again, the terms in $e^{-\lambda t}$ vanishes for large t, and we are left with 3 ω-periodic terms.

12.2 Linear systems – Our case

12.2.1 General computations, case of 3 cars

We consider the case of 3 cars. Let

$$X = \begin{pmatrix} u_2 \\ u_3 \end{pmatrix}$$

Then the system can be written as

$$X' = \begin{pmatrix} -\lambda & 0 \\ \lambda & -\lambda \end{pmatrix} X + \begin{pmatrix} \lambda u_1(t) \\ 0 \end{pmatrix}$$

which we write for short as $X' = AX + B(t)$. The matrix A has the eigenvalue $-\lambda$ with multiplicity 2. Its Jordan form can be obtained using the maple function `JordanForm`:

```
> with(LinearAlgebra)
> A := <<-lambda, lambda> | <0, -lambda>>:
> J := JordanForm(A)
```

giving

$$J = \begin{pmatrix} -\lambda & 1 \\ 0 & -\lambda \end{pmatrix}$$

To get the matrix of change of basis,

```
> P := JordanForm(A,output='Q')
```

giving

$$P = \begin{pmatrix} 0 & 1 \\ \lambda & 0 \end{pmatrix}$$

which is such that $P^{-1}AP = J$. Because $-\lambda$ is an eigenvalue with multiplicity 2 (the same multiplicity as as the size of the matrix), we can use the simplified theorem Theorem 9.4.13, and only need N. We have

$$\begin{aligned} N &= A - S \\ &= \begin{pmatrix} -\lambda & 0 \\ \lambda & -\lambda \end{pmatrix} - \begin{pmatrix} -\lambda & 0 \\ 0 & -\lambda \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ \lambda & 0 \end{pmatrix}. \end{aligned}$$

Clearly, $N^2 = 0$, so, by the theorem in the simplified case,

$$x(t) = e^{-\lambda t}\left(\mathbb{I} + Nt\right) x_0$$

But we know that solutions are unique, and that the solution to the differential equation is given by $x(t) = e^{At}x_0$. This means that

$$\begin{aligned} e^{At} &= e^{-\lambda t}\left(\mathbb{I} + Nt\right) \\ &= e^{-\lambda t}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \lambda t & 0 \end{pmatrix}\right) \\ &= e^{-\lambda t}\begin{pmatrix} 1 & 0 \\ \lambda t & 1 \end{pmatrix} \\ &= \begin{pmatrix} e^{-\lambda t} & 0 \\ \lambda t e^{-\lambda t} & e^{-\lambda t} \end{pmatrix}. \end{aligned}$$

Now notice that the solution to

$$X' = AX$$

is trivially established here, since

$$X(0) = \begin{pmatrix} u_2(0) \\ u_3(0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and thus

$$X(t) = e^{At}0 = 0.$$

The matrix e^{At} does however play a role in the solution (fortunately), since it is involved in the variation of constants formula:

$$X(t) = e^{At}X_0 + \int_0^t e^{A(t-s)}B(s)ds.$$

Thus we need to compute $e^{A(t-s)}B(s)$, and then the integral.

$$\begin{aligned} e^{A(t-s)}B(s) &= \begin{pmatrix} e^{-\lambda(t-s)} & 0 \\ \lambda(t-s)e^{-\lambda(t-s)} & e^{-\lambda(t-s)} \end{pmatrix} \begin{pmatrix} \lambda u_1(s) \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \lambda e^{-\lambda(t-s)}u_1(s) \\ \lambda^2 e^{-\lambda(t-s)}(t-s)u_1(s) \end{pmatrix} \end{aligned}$$

and thus

$$\begin{aligned} \int_0^t e^{A(t-s)}B(s)ds &= \int_0^t \begin{pmatrix} \lambda e^{-\lambda(t-s)}u_1(s) \\ \lambda^2 e^{-\lambda(t-s)}(t-s)u_1(s) \end{pmatrix} ds \\ &= \begin{pmatrix} \int_0^t \lambda e^{-\lambda(t-s)}u_1(s)ds \\ \int_0^t \lambda^2 e^{-\lambda(t-s)}(t-s)u_1(s)ds \end{pmatrix} \\ &= \begin{pmatrix} \lambda e^{-\lambda t}\int_0^t e^{\lambda s}u_1(s)ds \\ \lambda^2 e^{-\lambda t}\int_0^t e^{\lambda s}(t-s)u_1(s)ds \end{pmatrix} \\ &= \begin{pmatrix} \lambda e^{-\lambda t}\int_0^t e^{\lambda s}u_1(s)ds \\ \lambda^2 e^{-\lambda t}\left(t\int_0^t e^{\lambda s}u_1(s)ds - \int_0^t se^{\lambda s}u_1(s)ds\right) \end{pmatrix}. \end{aligned}$$

Let

$$\Psi(t) = \int_0^t e^{\lambda s}u_1(s)ds$$

and

$$\Phi(t) = \int_0^t se^{\lambda s}u_1(s)ds.$$

These can be computed when we choose a function $u_1(t)$. Then, finally, we have

$$\begin{aligned} X(t) &= \int_0^t e^{A(t-s)}B(s)ds \\ &= \begin{pmatrix} \lambda e^{-\lambda t}\Psi(t) \\ \lambda^2 e^{-\lambda t}\left(t\Psi(t) - \Phi(t)\right) \end{pmatrix}. \end{aligned}$$

12.2.2 Specialization to the case of the $\alpha \sin(\omega t)$ driver

We set

$$u_1(t) = \alpha \sin(\omega t).$$

Then

$$\Psi(t) = \frac{\alpha(\omega - \omega e^{\lambda t}\cos(\omega t) + \lambda e^{\lambda t}\sin(\omega t))}{\lambda^2 + \omega^2}$$

and

$$\Phi(t) = \frac{\alpha(\lambda^3 t + \lambda t\omega^2 - \lambda^2 + \omega^2)\sin(\omega t)e^{\lambda t}}{(\lambda^2+\omega^2)^2} - \frac{\alpha\omega\cos(\omega t)(t\lambda^2 + t\omega^2 - 2\lambda)e^{\lambda t} + 2\alpha\lambda\omega}{(\lambda^2+\omega^2)^2}$$

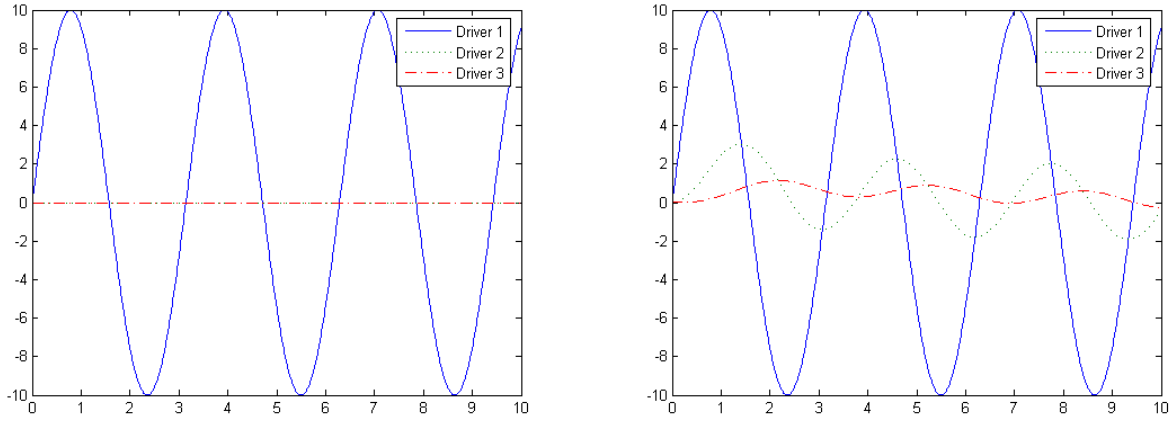


Figure 12.2: $\lambda = 0$ (left)and $\lambda = 0.4$ (right).

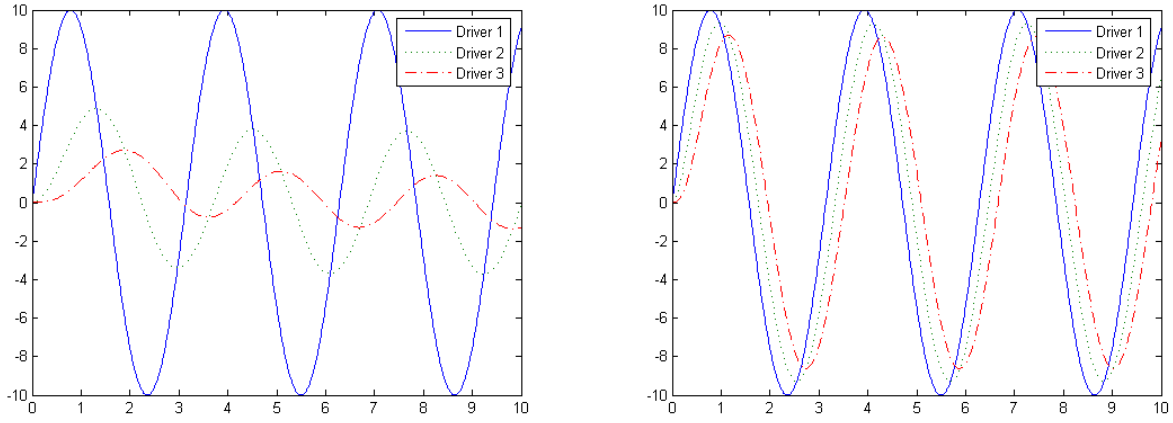


Figure 12.3: $\lambda = 0.8$ (left) and $\lambda = 5$ (right).

12.3 Prey-Predator model

1 prey:

$x(t)$ is the density of prey, and $y(t)$ is the density of predators

$$\begin{array}{ll} \frac{dx}{dt} & = x(r - \frac{r}{K}x - ay) \quad r, K, a > 0 \\ \frac{dy}{dt} & = y(-b + cx) \qquad\quad b, c > 0 \end{array}$$

- ay is the per capita loss of prey to the predator.
- cx is the per capita gain to the predator.

2 preys:

- x is the predator, it dies out in the absence of prey
- y is a prey, it grows exponentially in the absence of predators
- z is a prey, it grows logistically in the absence of predators

$$\begin{array}{rl} \frac{dx}{dt} = & \alpha xz + \beta xy - \gamma x \\ \frac{dy}{dt} = & \delta y - \epsilon xy \\ \frac{dz}{dt} = & \mu z(\nu - z) - \chi xz \end{array}$$

Equlibria:

$$(0,0,0) \quad and \quad (\frac{\delta}{\epsilon}, \frac{1}{\beta}(\gamma - \alpha\nu + \frac{\alpha\chi\delta}{\epsilon\mu}), \nu - \frac{\delta\chi}{\epsilon\mu})$$

12.4 Growth of living organisms

12.4.1 Michaelis-Menten enzyme kinetics

Michaelis Menten dynamics: e enzyme concentration, s subtrate concentration, c complexe concentration, p product concentration

$$\begin{aligned} \frac{ds}{dt} &= -k_1 se + k_{-1}c \\ \frac{de}{dt} &= -k_1 se + k_{-1}c + k_2 c \\ \frac{dc}{dt} &= k_1 se - k_{-1}c - k_2 c \\ \frac{dp}{dt} &= k_2 c \end{aligned}$$

- quasi-equilibrium hypothesis ($\frac{dc}{dt} = 0$) is valid when enzymes are efficient $e_0 << s_0$ (small concentration of enzymes in comparison to concentration of substrate).
- velocity of reaction $\frac{dp}{dt} = K_{max}\frac{s}{k_n + s}$, uptake in nutrient ($\frac{dS}{dt} = -K_{max}\frac{s}{k_n + s}$ where $K_{max} = k_2 e_0$ and $k_n = \frac{k_{-1}+k_2}{k_1}$),
- cooperative enzymes, Hill equation (generalization for n-subtrate complexes, $\frac{ds}{dt} = -K_{max}\frac{s^n}{k_n + s^n}$)

Chemostat model

Figure 12.4: Cell with unbounded transmembranar receptor x_0, bounded transmembranar receptors x_1. Nutrient n and p product.

- Michaelis-Menten dynamics can be used to describe the growth of bacteria from a given uptake of substrate.
- Inflow and outflow at a constant rate D to keep a constant volume in the chemostat with a concentration n_0 of subtrate in the inflow.

12.4.2 Kinetic reactions

Autocatalysis

12.4.3 Bifurcation

- Saddle-node bifurcation:
$$\frac{dx}{dt} = \mu - x^2$$
- Transcritical bifurcation:
$$\frac{dx}{dt} = \mu x - x^2$$
- Pitchfork bifurcation:
$$\frac{dx}{dt} = \mu x - x^3$$

Part V

Delay differential equations

Chapter 13

A brief theory of delay differential equations

We encountered a delay differential equation in Chapter 3, when we briefly discussed the delayed logistic equation (3.5). Here, we give more detail about this type of equations. The general theory of delayed differential equations is however out of the scope of these notes, so we remain superficial in our exposition and in the mathematical details.

13.1 Formulation of the problem

There are many types of delay differential equations (also called differential difference equations, or equations with deviating arguments). Our focus here will be on the simplest type, namely one for which the differential equation takes the form, for a given $\tau > 0$,

$$x' = f(x, x(t-\tau)), \tag{13.1}$$

and the corresponding initial value problem is given by

$$\begin{aligned} x' &= f(x, x(t-\tau)) \\ x(t) &= \phi(t), \quad \forall t \in [-\tau, 0], \end{aligned} \tag{13.2}$$

for some function $\phi \in C^0([-\tau, 0])$ called the **initial data**. Such an equation is called an autonomous equation (or system, if $x \in \mathbb{R}^n$) with **fixed time delay** τ. In (13.2), f depends on x at time t as well as x at time $t-\tau$. In the limit, if $\tau = 0$, then (13.2) reduces to an ordinary differential equation (this is a property that is used from time to time when analyzing a delay equation).

Other types of delay equations are, for example, the nonautonomous version of (13.1),

$$x' = f(t, x, x(t-\tau)),$$

equations with *distributed delay*,

$$x' = f\left(t, x, \int_{-\infty}^{t} x(t-s)ds\right)$$

and equations with *state dependent delay*,

$$x' = f(x, x(t-\tau(x(t)))).$$

13.2 Construction of the solution – The method of steps

The **method of steps** consists in considering (13.2) as a nondelayed IVP on the interval $[0, \tau]$. Indeed, on this interval, we can consider the IVP

$$\begin{aligned} x' &= f(x(t), \phi(t-\tau)) \\ x(0) &= \phi(0), \quad 0 \leq t \leq \tau. \end{aligned} \tag{13.3}$$

That the latter is a nondelayed problem is obvious if we rewrite the differential equation as

$$x'(t) = g(t, x(t)) \tag{13.4}$$

with

$$g(t, x(t)) = f(x(t), \phi(t-\tau)),$$

which is well defined on the interval $[0, \tau]$ since for $t \in [0, \tau]$, $t - \tau \in [-\tau, 0]$, on which the function ϕ is defined.

We can then use the integral form of the solution to construct the solution on the interval $[0, \tau]$,

$$\begin{aligned} x(t) &= x(0) + \int_0^t g(s, x(s))ds \\ &= \phi(0) + \int_0^t f(s, x(s), \phi(s-\tau))ds. \end{aligned}$$

Obviously, the nature of the solution depends on the function f. As problem (13.3) is an ordinary differential equations initial value problem, existence and uniqueness of solutions on the interval $[0, \tau]$ follow the usual scheme. To discuss the required properties on f and ϕ, the best is to use (13.4). Recall that a vector field has to be continuous both in t and in x for solutions to exist. Thus to have existence of solutions to the equation (13.4), g must be continuous in t and x. This implies that $f(x, \phi(t-\tau))$ must be continuous in t, x. Thus ϕ has to be continuous on $[-\tau, 0]$.

Now, for uniqueness of solutions to (13.4), we need g to be Lipschitz in x, *i.e.*, we require the same property from f. Note that this does not imply either ϕ_0 or the way f depends on ϕ_0.

Finally, remark that every successive integration raises the regularity of the solution: x is C^1 on $[0, \tau]$, C^2 on $[\tau, 2\tau]$, etc. Hence, x is C^n on $[(n-1)\tau, n\tau]$.

13.2.1 An example

Consider the delay initial value problem

$$\begin{aligned} x'(t) &= ax(t-\tau) \\ x(t) &= C, \quad t \in [-\tau, 0] \end{aligned} \tag{13.5}$$

with $a, C \in \mathbb{R}$, $\tau \in \mathbb{R}_+^*$. Using the method of steps, we want to find the solution to (13.5) on the interval $[k\tau, (k+1)\tau]$, $k \in \mathbb{N}$. We proceed as previously explained. To find the solution on the interval $[0, \tau]$, we consider the nondelayed IVP

$$\begin{aligned} x_1' &= ax_0(t) \\ x_1(0) &= C, \end{aligned}$$

where $x_0(t) = C$ for $t \in [0, \tau]$. The solution to this IVP is straightforward, $x_1(t) = C + aCt = C(1 + at)$, defined on the interval $[0, \tau]$. To integrate on the second interval, we consider the IVP

$$\begin{aligned} x_2' &= a[C(1 + at)] \\ x_2(\tau) &= x_1(\tau) = C + aC\tau. \end{aligned}$$

Hence we find the solution to the differential equation to be, on the interval $[\tau, 2\tau]$,

$$x_2(t) = C\left(1 + at + \frac{1}{2}a^2t^2 - \frac{1}{2}a^2\tau^2\right).$$

Iterating this process one more time with the IVP

$$\begin{aligned} x_3' &= a\left[C\left(1 + at + \frac{1}{2}a^2t^2 - \frac{1}{2}a^2\tau^2\right)\right] \\ x_3(2\tau) &= x_2(2\tau) = \frac{3}{2}a^2C\tau^2 + 2aC\tau + C. \end{aligned}$$

we find, on the interval $[2\tau, 3\tau]$, the solution

$$x_3(t) = C\left(1 + at + \frac{1}{2}a^2t^2 + \frac{1}{6}a^3t^3 - \frac{1}{2}ta^3\tau^2 - \frac{1}{3}a^3\tau^3 - \frac{1}{2}a^2\tau^2\right).$$

We develop the intuition that the solution at step n (*i.e.*, on the interval $[(n-1)\tau, n\tau]$) must take the form

$$x_n(t) = C\sum_{k=0}^{n} a^k \frac{(t - (k-1)\tau)^k}{k!}. \tag{13.6}$$

To show this, we proceed by induction. First, we check that it holds for $n = 1$, *i.e.*, $x_1(t) = C(1 + at)$ on the interval $[0, \tau]$. Now suppose that (13.6) holds for $n = j$, *i.e.*,

$$x_j(t) = C\sum_{k=0}^{j} a^k \frac{(t - (k-1)\tau)^k}{k!}$$

on the interval $[(j-1)\tau, j\tau]$. Then, we find the solution for the $(j+1)$th step by considering the nondelayed initial value problem,

$$\begin{aligned} x_{j+1}'(t) &= ax_j(t) \\ x_{j+1}(j\tau) &= x_j(j\tau) \end{aligned}$$

This is equivalent to

$$\begin{aligned} x_{j+1}'(t) &= aC\sum_{k=0}^{j} a^k \frac{(t - (k-1)\tau)^k}{k!} \\ x_{j+1}(j\tau) &= C\sum_{k=0}^{j} a^k \frac{((j\tau) - (k-1)\tau)^k}{k!} \end{aligned}$$

First, let us solve the differential equation. As the right hand side has an explicit form that does not involve x, it suffices to integrate the equation with respect to t. This gives, for $t \in [j\tau, (j+1)\tau]$,

$$\begin{aligned}
x_{j+1}(t) &= \int_{j\tau}^{t} aC \sum_{k=0}^{j} a^k \frac{(s-(k-1)\tau)^k}{k!} ds \\
&= aC \sum_{k=0}^{j} \int_{j\tau}^{t} a^k \frac{(s-(k-1)\tau)^k}{k!} ds \\
&= C \sum_{k=0}^{j} \left[a^{k+1} \frac{(s-(k-1)\tau)^{k+1}}{k!(k+1)} \right]_{s=j\tau}^{s=t} \\
&= C \sum_{k=0}^{j} a^{k+1} \left(\frac{(t-(k-1)\tau)^{k+1}}{(k+1)!} - \frac{(j\tau-(k-1)\tau)^{k+1}}{(k+1)!} \right) \\
&= C \sum_{k=0}^{j+1} a^{k+1} \left(\frac{(t-(k-1)\tau)^{k+1}}{(k+1)!} - \frac{((j-k+1)\tau)^{k+1}}{(k+1)!} \right)
\end{aligned}$$

13.2.2 Consequences of the method of steps

From the method of steps, it is clear that the initial data $\phi(t)$ must be continuous on the interval $[-\tau, 0]$. This is very different from ordinary differential equations, for which the initial condition is a single point.

The main consequence of this is that a delay differential equation is an object living in infinite dimensional space (the space of continuous functions) rather that in $\mathbb{R}^n$ as does a system of n ordinary differential equations.

Another interesting property that is the consequence of the method of steps is the increase of regularity of solutions with time.

Chapter 14

A delayed model of traffic flow

14.1 A delayed model of traffic flow

In the traffic flow model (12.1) of Chapter 12, reaction time is instantaneous. In practice, this is known to be incorrect: reflexes and psychology play a role. It takes at least a few instants to acknowledge a change of speed in the car in front. If the change of speed is not threatening, then you may not want to react right away. When you press the accelerator or the brake, there is a delay between the action and the reaction..

We consider the same setting as for system (12.1), except that now, for $t > 0$,

$$u'_{n+1}(t) = \lambda(u_n(t-\tau) - u_{n+1}(t-\tau)), \tag{14.1}$$

for $n = 1, \ldots, N-1$. Here, $\tau \geq 0$ is called the *time delay* (or *time lag*), or for short, *delay* (or *lag*). If $\tau = 0$, we are back to the model (12.1).

Initial data For a delay equation such as (14.1), initial data must be specified on an interval of length $N\tau$, left of zero.

This is easy to see by looking at the terms: $u(t-\tau)$ involves, at time t, the state of u at time $t-\tau$. So if $t < \tau$, we need to know what happened for $t \in [-\tau, 0]$. So, normally, we specify initial data as

$$u_n(t) = \phi(t) \text{ for } t \in [-\tau, 0],$$

where ϕ is some function, that we assume to be continuous. We assume $u_1(t)$ is known. Here, we assume, for $n = 1, \ldots, N$,

$$u_n(t) = 0, \qquad t \leq (n-1)\tau. \tag{14.2}$$

The explanation for this form of the initial data is simple. Since it takes τ units of time for each driver to adjust to changes in the speed of the car in front of them, car number 1 has to have moved for τ units of time before car number 2 makes any adjustment. In turn, car number 2 must have moved for τ units of time before car number 3 makes any adjustment. Repeating this, we obtain the form (14.2).

Important remark Although (14.1) looks very similar to (12.1), you must keep in mind that it is in fact much more complicated.

- A solution to (12.1) is a continuous function from $\mathbb{R}$ to $\mathbb{R}$ (or to $\mathbb{R}^n$ if we consider the system).
- A solution to (14.1) is a continuous function in the space of continuous functions.
- The space $\mathbb{R}^n$ has dimension n. The space of continuous functions has dimension ∞.

We can use the Laplace transform to get some understanding of the nature of the solutions.

14.2 Laplace transform of the DDE traffic flow model

Let

$$U_{k+1}(s) = \mathcal{L}\{u_{k+1}(t)\} = \int_0^\infty e^{-st} u_{k+1}(t)dt.$$

Since we have assumed initial data of the form

$$u_n(t) = 0 \qquad \text{for } t \leq (n-1)\tau,$$

we have

$$U_{k+1}(s) = \int_{k\tau}^\infty e^{-st} u_{k+1}(t)ds.$$

Since $u_{n+1}(t) = 0$ for $t \leq n\tau$,

$$\begin{aligned}\int_0^\infty e^{-st} u'_{n+1}(t)dt &= \left[u_{k+1}(t)e^{-st}\right]_{k\tau}^\infty + s\int_{k\tau}^\infty e^{-st}u_{k+1}(t)dt \\ &= sU_{k+1}(s)\end{aligned}$$

and

$$\begin{aligned}\int_0^\infty e^{-st}u_{k+1}(t-\tau)dt &= \int_{(k-1)\tau}^\infty e^{-st}u_{k+1}(t-\tau)dt \\ &= \int_{(k-2)\tau}^\infty e^{-s(t+\tau)}u_k(\tau)d\tau \\ &= e^{-s\tau}U_k(s),\end{aligned}$$

since $e^{-st}u_{k+1}(t) \to 0$ for the improper integral to exist. Note that we could have obtained this directly using the properties of the Laplace transform.

Multiply

$$u'_{n+1}(t) = \lambda(u_n(t-\tau) - u_{n+1}(t-\tau))$$

by e^{-st},

$$e^{-st}u'_{n+1}(t) = \lambda e^{-st}(u_n(t-\tau) - u_{n+1}(t-\tau))$$

integrate over $(0,\infty)$ (using the expressions found above),

$$sU_{n+1}(s) = \lambda(e^{-s\tau}U_n(s) - e^{-s\tau}U_{n+1}(s))$$

which is equivalent to

$$U_{n+1}(s) = \frac{\lambda U_n(s)}{\lambda + se^{s\tau}}$$

Thus, when $U_1(s)$ is known, we can deduce the values for all U_n. Suppose

$$u_1(t) = \alpha \sin(\omega t)$$

From the table of Laplace transforms, it follows that

$$U_1(s) = \alpha \frac{\omega}{s^2 + \omega^2}$$

Therefore,

$$U_2 = \frac{\lambda U_1(s)}{\lambda + se^{st}} = \alpha \frac{\lambda}{\lambda + se^{st}} \frac{\omega}{s^2 + \omega^2}$$

and we can continue..

However, even though we know the solution in s-space, it is difficult to get the behavior in t-space, by hand, and maple does not help us either.

Part VI

Partial differential equations

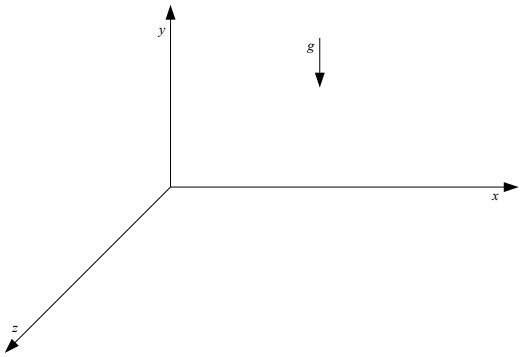
Chapter 15

Shallow water

Partial differential equations

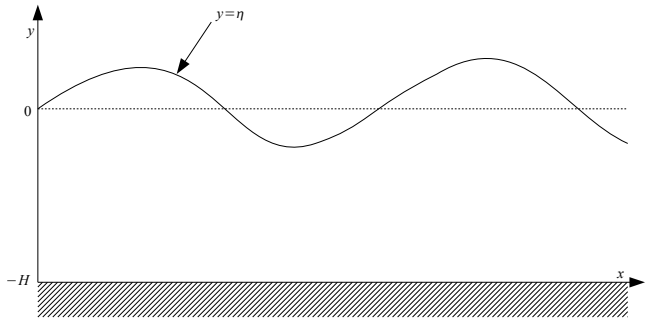
15.1 Model formulation

Spatial domain We consider the motion of a body of water that is infinite in the z direction, with or without boundary in the x direction, and the vertical direction of gravity taken as the y direction.



From now on, suppose z direction uniform (the same for all z), so ignore z except for the sake of argument.

- Water depth at rest, H, small compared to distance L_0 over which significant changes can occur in the x direction.
- Undisturbed water surface, $y = 0$.
- Moving upper free surface $y = \eta$, measured from $y = 0$.
- Sea floor $y = -H$.

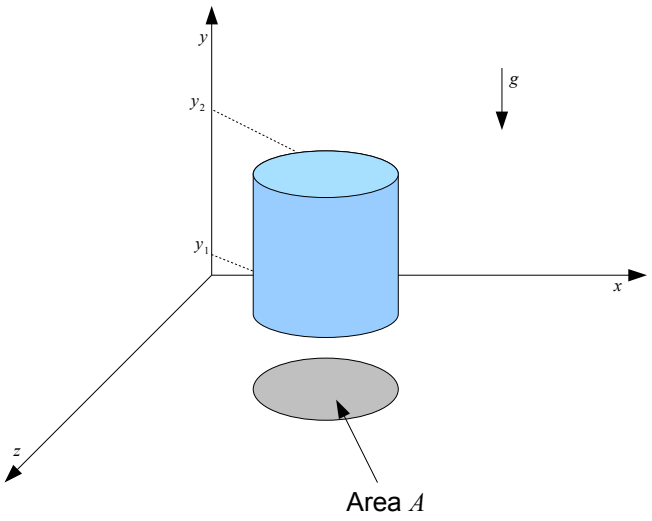


- u velocity in the x direction. Assume independent of depth y.
- ρ mass density of water.
- $p(x, y, t)$ pressure in fluid at point (x, y) at time t. In water, magnitude at any (x, y) is same in all directions.

Fluid motion independent of z, so

- $u = u(x, t)$
- $\eta = \eta(x, t)$.

Take a cylindrical water column, with base area A, between y_1 and $y_2 > y_1$.



Force equilibrium in the y direction in this cylinder requires balance of weight of water column and pressure differential between bottom face $y = y_1$ and top face $y = y_2$.

Weight of water column:

$$\iint_A \int_{y_1}^{y_2} (-\rho g)\, dydxdz$$

Pressure differential:

$$\iint_A (p(x, y_2, t) - p(x, y_1, t))\ dxdz$$

So we must have

$$\iint_A \int_{y_1}^{y_2} (-\rho g)\ dydxdz = \iint_A (p(x, y_2, t) - p(x, y_1, t))\ dxdz$$

$$\iint_A \int_{y_1}^{y_2} (-\rho g)\ dydxdz = \iint_A (p(x, y_2, t) - p(x, y_1, t))\ dxdz$$

is equivalent to

$$\iint_A \int_{y_1}^{y_2} \left(\frac{\partial p}{\partial y} + \rho g\right)\ dydxdz = 0$$

This must be true for any water column, i.e., any A, y_1, y_2. Therefore,

$$\frac{\partial p}{\partial y} + \rho g = 0$$

(otherwise, we would be able to find a water column where the integrand is positive, leading to a positive value of the integral on that column).

Water is incompressible If you force a body of water to deform, the volume of that body of water remains constant, i.e., water is an *incompressible fluid*.

$\Rightarrow \rho$, the density, is a constant, and from

$$\frac{\partial p}{\partial y} + \rho g = 0$$

we get

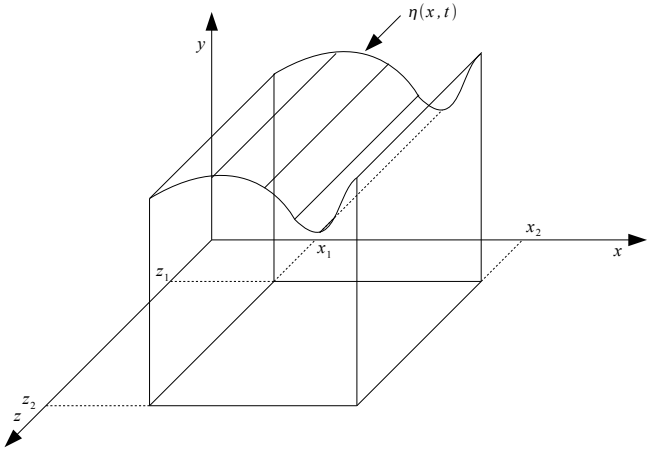
$$p = -\rho g y + C,$$

so if p is measured relative to the pressure above the free upper surface $y = \eta$,

$$p = \rho g(\eta - y)$$

Water accumulation Consider a fixed volume V,

$$V = \{z_1 \leq z \leq z_2, x_1 \leq x \leq x_2, -H \leq y \leq \eta\}$$



Water enters V through x_1 face and leaves V through x_2 face. Rate of water accumulation in V is

$$\frac{d}{dt}\int_{z_1}^{z_2}\int_{x_1}^{x_2}\int_{-H}^{\eta}\rho\, dydxdz = \Delta z\frac{d}{dt}\int_{x_1}^{x_2}\rho h\, dx,$$

with $\Delta z = z_2 - z_1$, and $h(x,t) = \eta + H$ the height of water at time t at spatial location x.

Water flux Net flux of water entering V through its faces $x = x_1$ and $x = x_2$ is

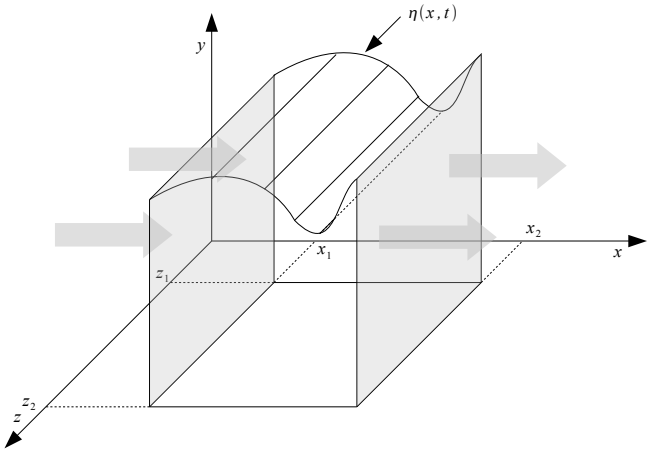


Figure 15.1: Flows through the domain.

$$\left[\int_{z_1}^{z_2}\int_{-H}^{\eta} u\, dydz\right]_{x=x_1} - \left[\int_{z_1}^{z_2}\int_{-H}^{\eta} u\, dydz\right]_{x=x_2} = -\Delta z\left[\rho uh\right]_{x_1}^{x_2}$$

There is no flux through $y = -H$ and $y = \eta$, and no net flux through $z = z_1$ and $z = z_2$.

Conservation of mass Of course, the mass must conserve in V, so the two expressions must be equal, i.e.,

$$\frac{d}{dt}\int_{x_1}^{x_2}\rho h\, dx + \left[\rho uh\right]_{x_1}^{x_2} = 0$$

Newton's second law for deformable media (Euler): rate of increase of horizontal momentum (in the x direction) in V must equal the sum of the net influx of momentum into the volume and the net horizontal force acting on the column. (Momentum: product of mass and velocity of an object).

Rate of increase of momentum

$$\frac{d}{dt}\int_{z_1}^{z_2}\int_{x_1}^{x_2}\int_{-H}^{\eta}\rho u\, dydxdz = \Delta z\frac{d}{dt}\int_{x_1}^{x_2}\rho uhdx$$

Momentum flux Net influx of momentum through faces $x = x_1$ and $x = x_2$ is

$$\left[\int_{z_1}^{z_2}\int_{-H}^{\eta}(\rho u)u\ dydz\right]_{x=x_1} - \left[\int_{z_1}^{z_2}\int_{-H}^{\eta}(\rho u)u\ dydz\right]_{x=x_2} = -\Delta z\left[\rho u^2 h\right]_{x_1}^{x_2}.$$

There is no flux through $y = -H$ and $y = \eta$, and no net flux through $z = z_1$ and $z = z_2$.

Forces acting on V Ignore friction at $y = -H$. Then only contributions to horizontal forces come from pressure at $x = x_1$ and $x = x_2$, so net horizontal forces acting on V is

$$\begin{aligned}\left[\int_{z_1}^{z_2}\int_{-H}^{\eta} p\ dydz\right]_{x_1}^{x_2} &= -\left[\Delta z\int_{-H}^{\eta}\rho g(\eta - y)\ dy\right]_{x_1}^{x_2}\\ &= \left[-\Delta z\rho g(\eta y - \frac{1}{2}y^2)\Big|_{-H}^{\eta}\right]_{x_1}^{x_2}\\ &= \left[-\frac{1}{2}\Delta z\rho g h^2\right]_{x_1}^{x_2}\end{aligned}$$

Conclusion from Newton's second law

$$\frac{d}{dt}\int_{x_1}^{x_2}\rho u h\ dx + \left[\rho u^2 h + \frac{1}{2}\rho g h^2\right]_{x_1}^{x_2} = 0$$

The general model Pressure magnitude:

$$p = \rho g(\eta - y) \tag{15.1}$$

Horizontal velocity:

$$\frac{d}{dt}\int_{x_1}^{x_2}\rho h\ dx + [\rho u h]_{x_1}^{x_2} = 0 \tag{15.2}$$

Free surface height:

$$\frac{d}{dt}\int_{x_1}^{x_2}\rho u h\ dx + \left[\rho u^2 h + \frac{1}{2}\rho g h^2\right]_{x_1}^{x_2} = 0 \tag{15.3}$$

15.2 Case of smooth solutions

Suppose u and h are smooth (with continuous first order partial derivatives), then (15.2) and (15.3) take a much simpler form,

$$\int_{x_1}^{x_2}\left(\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(uh)\right)\ dx = 0$$

and

$$\int_{x_1}^{x_2} \left(\frac{\partial}{\partial t}(uh) + \frac{\partial}{\partial x}(u^2 h + \frac{1}{2} g h^2) \right) \, dx = 0$$

Since the intervals of integration $[x_1, x_2]$ are arbitrary, and that the integrands are continuous, we have

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(uh) = 0$$

and

$$\frac{\partial}{\partial t}(uh) + \frac{\partial}{\partial x}(u^2 h + \frac{1}{2} g h^2) = 0$$

We write

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(uh) = 0$$

and

$$\frac{\partial}{\partial t}(uh) + \frac{\partial}{\partial x}(u^2 h + \frac{1}{2} g h^2) = 0$$

as

$$h_t + (uh)_x = 0 \tag{15.4}$$

and

$$(uh)_t + (u^2 h + \frac{1}{2} g h^2)_x = 0 \tag{15.5}$$

From (15.4),

$$h_t = -(uh)_x = -(u_x h + u h_x)$$

Equation (15.5) can be rewritten as

$$\begin{aligned}
(15.5) &\Leftrightarrow u_t h + u h_t + (u^2 h + \frac{1}{2} g h^2)_x = 0 \\
&\Leftrightarrow u_t h - u(u_x h + u h_x) + 2 u u_x h + u^2 h_x + g h h_x = 0 \\
&\Leftrightarrow u_t h - u u_x h - \cancel{u^2 h_x} + 2 u u_x h + \cancel{u^2 h_x} + g h h_x = 0 \\
&\Leftrightarrow u_t h + u u_x h + g h h_x = 0
\end{aligned}$$

Therefore, provided $h \neq 0$, we get

$$h_t + (uh)_x = 0 \tag{15.6a}$$

$$u_t + u u_x + g h_x = 0 \tag{15.6b}$$

which describes the evolution of u and h.

The model for smooth solutions

$$h_t + (uh)_x = 0 \tag{15.6a}$$

$$u_t + u u_x + g h_x = 0 \tag{15.6b}$$

If $-\infty < x < \infty$, then all we need is an initial condition, i.e., functions describing the initial state of u and h:

$$u(x, 0) = u_0(x), \quad h(x, 0) = h_0(x), \qquad -\infty < x < \infty.$$

If x has a boundary, then we need boundary conditions.

15.3 Linearization

Suppose the bottom is flat (H is constant), and that the deviation from the undisturbed depth H is small compared to H itself, then

$$h = (H + \zeta) = H(1 + \frac{\zeta}{H}) \simeq H, \qquad h_t = \zeta_t, \qquad h_x = \zeta_x.$$

If $|u|$ is also small, then uu_x can be neglected. Then we can linearize

$$h_t + (uh)_x = 0 \tag{15.6a}$$
$$u_t + uu_x + gh_x = 0, \tag{15.6b}$$

getting

$$\zeta_t + Hu_x = 0 \tag{15.7a}$$
$$u_t + g\zeta_x = 0 \tag{15.7b}$$

Differentiate (15.7b) with respect to x:

$$u_{tx} + g\zeta_{xx} = 0$$

and therefore,

$$u_{tx} = -g\zeta_{xx} \tag{15.8}$$

Differentiate (15.7a) with respect to t:

$$\zeta_{tt} + Hu_{xt} = 0 \tag{15.9}$$

If u has continuous second-order partial derivatives, then from Clairaut's theorem, $u_{tx} = u_{xt}$. Therefore, substituting (15.8) into (15.9),

$$\zeta_{tt} - HG\zeta_{xx} = 0$$

that is

$$\zeta_{tt} = c^2\zeta_{xx}, \qquad c^2 = Hg$$

The one-dimensional wave equation (1) The partial differential equation

$$\zeta_{tt} = c^2\zeta_{xx} \tag{15.10}$$

with $c^2 = Hg$, is the one-dimensional wave equation. Initial conditions are given by

$$\zeta(x, 0) = h_0(x) - H \equiv \zeta_0(x)$$
$$\zeta_t(x, 0) = -Hu_x(x, 0) = -H[u_0(x)]_x \equiv \nu_0(x)$$

The one-dimensional wave equation (2) Things can also be expressed in terms of u. Using the same type of simplification used before for ζ, we get

$$u_{tt} = c^2 u_{xx} \tag{15.11}$$

with $c^2 = Hg$. Initial conditions are given by

$$\begin{aligned} u(x,0) &= u_0(x) \\ u_t(x,0) &= -g\zeta_x(x,0) = -g[h_0(x)]_x \equiv v_0(x) \end{aligned}$$

15.4 Traveling wave solutions

Traveling wave solutions This was obtained by d'Alembert. Consider

$$u_{tt} = c^2 u_{xx} \tag{15.11}$$

Note that this can be written as

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) u = 0$$

This implies that for any F, G, the sum

$$u(x,t) = F(x - ct) + G(x + ct)$$

satisfies (15.11).

Derivation of the solution Introduce the new variables

$$a = x - ct \qquad \text{and} \qquad b = x + ct$$

We have

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial a} + \frac{\partial u}{\partial b} \qquad \frac{\partial u}{\partial t} = -c\frac{\partial u}{\partial a} + c\frac{\partial u}{\partial b}$$

$$\frac{\partial^2}{\partial x^2}u = \left(\frac{\partial}{\partial a} + \frac{\partial}{\partial b}\right)^2 u = \frac{\partial^2 u}{\partial a^2} + 2\frac{\partial^2 u}{\partial a \partial b} + \frac{\partial^2 u}{\partial b^2}$$

$$\frac{\partial^2}{\partial t^2}u = \left(-c\frac{\partial}{\partial a} + c\frac{\partial}{\partial b}\right)^2 u = c^2\left(\frac{\partial^2 u}{\partial a^2} - 2\frac{\partial^2 u}{\partial a \partial b} + \frac{\partial^2 u}{\partial b^2}\right)$$

So the equation

$$u_{tt} = c^2 u_{xx} \tag{15.11}$$

is written

$$4\frac{\partial^2 u}{\partial a \partial b} = 0$$

Integrate with respect to b:

$$\frac{\partial u}{\partial a} = \xi(a)$$

and thus

$$\begin{aligned} u(x,t) = u(a,b) &= \int \xi(a)da + G(b) \\ &= F(a) + G(b) \\ &= F(x-ct) + G(x+ct) \end{aligned}$$

Set

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x)$$

Then d'Alembert's formula gives

$$u(x,t) = \frac{f(x-ct) + f(x+ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)ds$$

Case of a Dirac delta initial condition Suppose $u_0(x) = 0$ and $v_0(x) = \delta(x)$, for $-\infty < x < \infty$, with δ the Dirac delta,

$$\delta(x) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

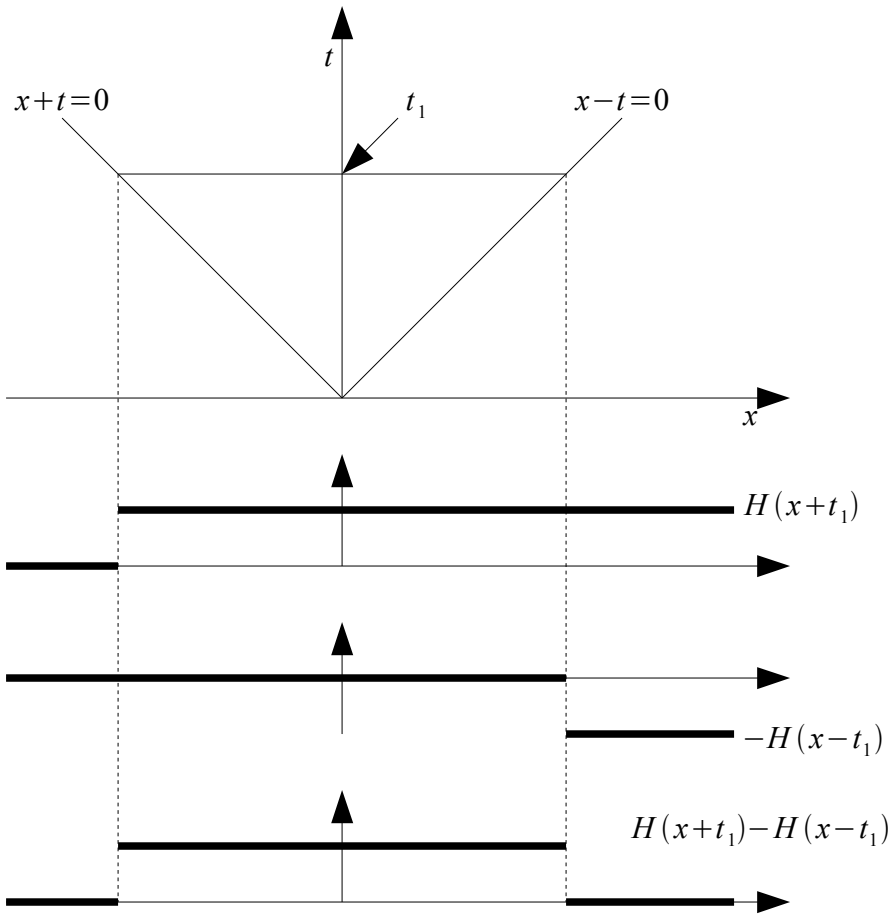
$$u(x,t) = \frac{1}{2c}\int_{x-ct}^{x+ct} \delta(z)dz = \frac{1}{2c}\left\{H(x+ct) - H(x-ct)\right\},$$

with H the Heaviside function,

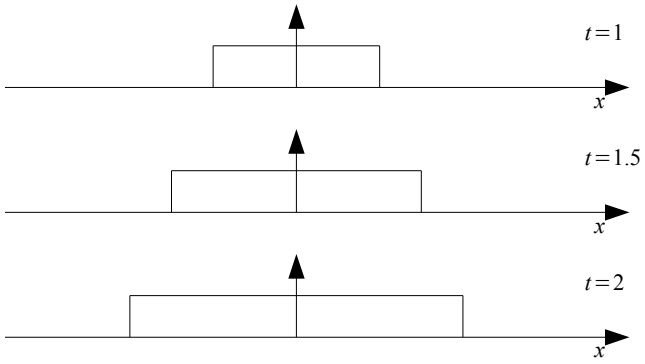
$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0. \end{cases}$$

For simplicity, take $c = 1$. This gives

$$u(x,t) = \frac{1}{2}\left\{H(x+t) - H(x-t)\right\},$$



As t increases, we move further up in the top graph in (x,t)-space, resulting in a wider and wider square pulse.



Appendix A

Descartes' rule of signs

Descartes' rule of signs is a very useful way to study the existence and sign of the roots of a polynomial without having to express them explicitely.

Theorem A.1 (Descartes' rule of signs)**.** *Let $p(x) = \sum_{i=0}^{m} a_i x^i$ be a polynomial with real coefficients such that $a_m \neq 0$. Define v to be the number of variations in sign of the sequence of coefficients $a_m, \ldots, a_0$. By 'variations in sign' we mean the number of values of n such that the sign of a_n differs from the sign of a_{n-1}, as n ranges from m down to 1. Then*

- *the number of positive real roots of $p(x)$ is $v - 2N$ for some integer N satisfying $0 \leq N \leq \dfrac{v}{2}$,*
- *the number of negative roots of $p(x)$ may be obtained by the same method by applying the rule of signs to $p(-x)$.*

Example – Let $p(x) = x^3 + 3x^2 - x - 3$. The coefficients have sign $++--$, so there is one sign change. Thus $v = 1$. Since $0 \leq N \leq 1/2$, we must have $N = 0$. Thus $v - 2N = 1$ and there is exactly one positive real root of $p(x)$.

To find the negative roots, we examine $p(-x) = -x^3 + 3x^2 + x - 3$. The coefficients have sign $-++-$, so there are two sign changes. Thus $v = 2$ and $0 \leq N \leq 2/2 = 1$. Thus, there are two possible solutions, $N = 0$ and $N = 1$, and two possible values of $v - 2N$. Therefore, there are either two or no negative real roots. Furthermore, note that $p(-1) = (-1)^3 + 3 \cdot (-1)^2 - (-1) - 3 = 0$, hence there is at least one negative root. Therefore there must be exactly two. ◇

Appendix B

Some matrix theory

B.1 Eigenvalues and eigenvectors

Let $M \in \mathcal{M}_n(\mathbb{F})$ with $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. The **eigenvalues** of M are numbers $\lambda \in \mathbb{C}$ found by solving the equation

$$\det(M - \lambda \mathbb{I}) = 0, \tag{B.1}$$

where $\mathbb{I}$ is the identity matrix of $\mathcal{M}_n(\mathbb{F})$, and $v \in \mathbb{F}^n$. Another way to write (B.1) is as

$$Mv = \lambda v.$$

It is easy to see that these two expressions are equivalent. If $M \in \mathcal{M}_n(\mathbb{R})$, then there are exactly n eigenvalues in $\mathbb{C}$ (or $\mathbb{R}$), including multiplicity. This set of values is called the **spectrum** of M, and is usually denoted Sp (M) or spec(M). In other words,

$$\text{Sp } (M) = \{\lambda \in \mathbb{C} : \det(M - \lambda \mathbb{I}) = 0 \text{ for some } v \in \mathbb{C}^n\}.$$

Note that eigenvalues are *matrix invariants*, in the sense that they are preserved by linear transformations of the vector space. (Other examples of matrix invariants include the rank, the determinant and the trace.) Another name for (B.1) is the **characteristic polynomial**, which is obtained by considering the polynomial resulting from (B.1),

$$P(\lambda) = \det(M - \lambda \mathbb{I}).$$

Eigenvalues of M are then the roots of $P(\lambda)$.

To a given eigenvalue $\lambda_i \in \text{Sp } (M)$, there corresponds an **eigenvector** v_i which satisfies the equation (B.1) for $\lambda = \lambda_i$.

B.1.1 Left eigenvectors

Let M be an $r \times r$ matrix, u, v be two column vectors, $\lambda \in \mathbb{R}$. Then, if

$$Mu = \lambda u,$$

u is the (right) eigenvector corresponding to λ, and if

$$v^T M = \lambda v^T$$

then v is the left eigenvector corresponding to λ. Note that to a given eigenvalue there corresponds (to a multiple) one left and one right eigenvector.

B.2 Tools to determine properties of eigenvalues

Theorem B.1. *(Routh-Hurwitz Criteria) Given the polynomial,*

$$P(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$$

where the coefficients a_i are real constants, $i = 1, \ldots, n$ define the n Hurwitz matrices using the coefficients a_i of the characteristic polynomial:

$$H_1 = (a_1), \quad H_2 = \begin{pmatrix} a_1 & 1 \\ a_3 & a_2 \end{pmatrix}, \quad H_3 = \begin{pmatrix} a_1 & 1 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{pmatrix},$$

and

$$H_n = \begin{pmatrix} a_1 & 1 & 0 & 0 & \ldots & 0 \\ a_3 & a_2 & a_1 & 1 & \ldots & 0 \\ a_5 & a_4 & a_3 & a_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & a_n \end{pmatrix}$$

where $a_j = 0$ if $j > n$. All of the roots of the polynomial $P(\lambda)$ are negative or have negative real part if and only if the determinants of all Hurwitz matrices are positive:

$$\det H_i > 0, \quad j = 1, 2, \ldots, n.$$

Theorem B.2. *(Corollary) Routh-Hurwitz criteria for $n = 2, 3, 4, 5$*

- $n = 2 : a_1 > 0$ *and* $a_2 > 0$.
- $n = 3 : a_1 > 0$, $a_3 > 0$ *and* $a_1a_2 > a_3$.
- $n = 4 : a_1 > 0$, $a_3 > 0$, $a_4 > 0$ *and* $a_1a_2a_3 > a_3^2 + a_1^2a_4$.
- $n = 5 : a_i > 0$, $i = 1, 2, 3, 4, 5$, $a_1a_2a_3 > a_3^2 + a_1^2a_4$ *and* $(a_1a_4 - a_5)(a_1a_2a_3 - a_3^2 - a_1^2a_4) > a_5(a_1a_2 - a_3)^2 + a_1a_5^2$

Theorem B.3. *(Gerhgorin's Theorem) Let A be an $n \times n$ matrix. Let D_i be the disk in the complex plane with the center at a_{ii} and radius $r_i = \sum_{j=1, j\neq i}^{n} |a_{ij}|$. Then all eigenvalues of the matrix A lie in the union of the disks D_i, $i = 1, 2, \ldots, n$, $\cup_{i=1}^{n} D_i$. In particular, if λ is an eigenvalue of A, then for some $i = 1, 2, \ldots, n$*

$$|\lambda - a_{ii}| \leq r_i.$$

Theorem B.4. *(Corollary) Let A be an $n \times n$ matrix with real entries. If the diagonal elements of A satisfy*

$$a_{ii} < -r_i \quad \textit{where} \quad r_i = \sum_{j=1, j\neq i}^{n} |a_{ij}|$$

for $i = 1, 2, \ldots, n$ then the eigenvalues of A are negative or have have negative real part.

Definition B.2.1. *The spectral radius of matrix A is denoted $\rho(A)$ and is defined as*

$$\rho(A) = \max_{i \in \{1, 2, \ldots, m\}} |\lambda_i|$$

Theorem B.2.2. *Let A be a $k \times k$ matrix*

$$\rho(A) \leq \|A\|$$

Three usual matrix norms are

- $\|A\|_1 = \max_{1\leq j\leq k} \sum_{i=1}^{k} |a_{ij}|$ (sum over columns),
- $\|A\|_2 = [\rho(A^T A)]^{1/2}$
- $\|A\|_\infty = \max_{1\leq i\leq k} \sum_{j=1}^{k} |a_{ij}|$ (sum over rows).

Theorem B.2.3. *Let A be a constant $m \times m$ matrix. Then the spectral radius of A satisfies $\rho(A) < 1$ if and only if*

$$\lim_{t\to+\infty} A^t = \mathbf{0}$$

B.3 Nonnegative matrices

Definition B.3.1. *A matrix A whose entries are nonnegative is called a **nonnegative matrix**, denoted $A \geq 0$.*

Definition B.3.2. *A matrix A whose entries are positive is called a **positive matrix**, denoted $A > 0$.*

Definition B.3.3. *A square $m \times m$ matrix $A = a_{ij}$ is **reducible** if the index set $1, 2, \ldots, m$ can be split into two nonempty complementary sets S_1 and S_2: $S_1 = \{i_1, \ldots, i_\mu\}$ and $S_2 = \{k_1, \ldots, k_\varepsilon\}$ where $m = \mu + \varepsilon$ such that*

$$a_{i_\alpha k_\beta} = 0 \quad (\alpha = 1, 2, \ldots, \mu; \beta = 1, 2, \ldots, \varepsilon).$$

*Otherwise, the matrix A is **irreducible**.*

Definition B.3.4. *If there exits a directed path from node i to j for every node i and j in the digraph, then the digraph is said to be **strongly connected**.*

Theorem B.3.5. *The digraph of matrix A is strongly connected if and only if A is irreducible.*

Theorem B.3.6 (Frobenius theorem)**.** *An irreducible, nonnegative matrix A always has a positive eigenvalues λ that is a simple root (multiplicity one) of the characteristic equation. The value of λ is greater than or equal to the magnitude of all the other eigenvalues. To the eignevalue λ, there corresponds an eigenvector with positive coordinates.*

Theorem B.3.7 (Perron theorem)**.** *A positive matrix A always has a real, positive eigenvalue λ that is a simple root of the characteristic equation and exceeds the magnitude of all of the other eigenvalues*

$$|\lambda_i| < |\lambda|, \quad \forall i.$$

To the eigenvalue λ there corresponds an eigenvector with positive coordinates.

In other words, $\rho(A)$ is a positive real eigenvalue of A with multiplicity 1.

Definition B.3.8 (Primitivity)**.** *If an irreducible, nonnegative matrix A has h eigenvalues $\lambda_1, \lambda_2, \dots \lambda_h$ of maximum modulus ($|\lambda_1| = |\lambda_i|, i = 1, 2, \dots, h$), then A is called* ***primitive*** *if $h = 1$ and* ***imprimitive*** *if $h > 1$. The value of h is called the* ***index of imprimitivity****.*

The index of imprimitivity is the number of eigenvalues of matrix A with maximum modulus (with magnitude equal to $\rho(A)$).

Theorem B.3.9. *A nonnegative matrix A is primitive if and only if some power of A is positive (i.e. $A^p > 0$ for some integer $p \geq 1$).*

Theorem B.3.10. *[3] A irreducible matrix is primitive if its trace if positive.*

The following theorem is one of the most important in the theory of matrices.

Theorem B.3.11. *(Perron-Frobenius Theorem) If M is a nonnegative primitive matrix, then:*

- *M has a positive eigenvalue λ_1 of maximum modulus.*
- *λ_1 is a simple root of the characteristic polynomial.*
- *for every other eigenvalue λ_i, $\lambda_1 > \lambda_i$ (it is strictly dominant)*
- $$\min_i \sum_j m_{ij} \leq \lambda_1 \leq \max_i \sum_j m_{ij}$$
  $$\min_j \sum_i m_{ij} \leq \lambda_1 \leq \max_j \sum_i m_{ij}$$
- *the row and column eigenvectors associated with λ_1 are strictly positive.*
- *the sequence M^t is asymptotically one-dimensional, its columns converge to the column eigenvector associated with λ_1; and its rows converges to the row eigenvector associated with λ_1.*

B.3.1 Suggested reading

Berman and Plemmons [2, 3] provides a very nice description of the links between nonnegative matrices and graphs. Senetta..

Bibliography

[1] L.J.S. Allen. *An Introduction to Mathematical Biology*. Pearson, 2007.

[2] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979.

[3] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics. SIAM, 1994.

[4] S.N. Elaydi and W.A.Jr Harris. On the computation of a^n. *SIAM Review*, 40:965–971, 1998.

[5] J.N. Franklin. *Matrix Theory*. Dover, 2000.

[6] L. Glass and M. Mackey. *From clock to chaos*. Princeton University Press, 1988.

[7] G.E. Hutchinson. Circular causal systems in ecology. *Ann. N. Y. Acad. Sci.*, 50:221–246, 1948.

[8] W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. London, Ser. A*, 115:700–721, 1927.

[9] W.O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics. ii. the problem of endemicity. *Proc. Roy. Soc. London, Ser. A*, 138:55–83, 1932.

[10] W.O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics. iii. further studies of the problem of endemicity. *Proc. Roy. Soc. London, Ser. A*, 141:94–122, 1933.

[11] R. Pearl and L.J. Reed. On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Science*, 6:275–288, 1920.

[12] R. Pearl and L.J. Reed. The logistic curve and the census count of 1930. *Science*, 72(1868):399–401, 1930.

[13] R. Pearl, L.J. Reed, and J.F. Fish. The logistic curve and the census count of 1940. *Science*, 92(2395):486–488, 1940.

[14] E.C. Pielou. *Mathematical ecology*. John Wiley, New York, 1977.

[15] H.S. Pritchett. A formula for predicting the population of the united states. *Publications of the American Statistical Association*, 2(14):278–286, 1891.

[16] P.F. Verhulst. Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathematique et Physique*, 10:113–121, 1838.

[17] P.F. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouv. mém. de l'Academie Royale des Sci. et Belles-Lettres de Bruxelle*, 18:1–41, 1845.

[18] E.M. Wright. A non-linear difference-differential equation. *J. Reine Angew. Math.*, 494:66–87, 1955.

Index