# Rules for computer lab

## General Instructions

The information in this laboratory should be solved individually or in pairs and summarized in a report. The report must be submitted via the Computer Computing (Tasks) tool in Cambro, in the form of a PDF. In order to facilitate the correction, the name(s) of the person(s) who wrote the report should be in the filename, for example "labreport_firstname_lastname.pdf". If you work in pairs, one of you will submit the report and state with whom you wrote the report.

The report must be submitted via Cambro no later than the 9th of February by 16:00. If the laboratory report is not satisfactory, a new report must be submitted by the 9th of March by 16:00. There is a final opportunity to submit a lab report by the 6th of April at 16:00. If the report still does not pass, there is the opportunity to do another computer lab at the next time the course is given, but it will then be according to the instructions applicable at the time of the course (i.e., then it may be a new computer lab with other tasks to be solved).

Supervision in the computer room is offered on the 1st of February, 2018. Come well-prepared; it is recommended that you read the instructions for the lab report carefully and do some planning regarding the computer code and the analysis beforehand. You can also use the campus license for MATLAB to do the computer lab at another time. Read more about running MATLAB on a private computer at: `http://www.math.umu.se/for-vara-studenter/matlab/`.

## Instructions for the report

The lab is separated into different conceptual sections. The first is background information that provides a context for the tasks you will perform. The other sections guide you through different analyses that ultimately allow you evaluate the validity of a scientific hypothesis and formulate a new one. The tasks require you to answer questions, offer explanations, write MATLAB code, and submit graphs.

**Written material** For the written material, please write at a level understandable to a third party who does not have access to the laboratory instructions, using complete sentences rather than sentence fragments. Assume that your audience has a basic grasp of linear algebra concepts such as vectors and matrices.

**Graphs** Graphs should have labels on the horizontal and vertical axes (use `xlabel` and `ylabel`). For this lab report, the `plot` command should be sufficient to generate graphs, so `plot(xdata,ydata,'o-')` should work well to plot ydata versus xdata.

**Computer code** Provide MATLAB code for tasks when asked. The code should be such that someone could copy and paste it into a brand new MATLAB prompt and it should work. In other words, if you type `clear` into MATLAB and then paste your code it should work.

**Numerical values** There are instances in this lab when you are asked to find the value of a variable $p$ that produces the most bacteria. To be correct, your value of $p$ must be within $\pm .025$ of the actual best value.

**Identification** Please clearly indicate your name, date, and the name of the course on your report.

## Correction protocol

At the end of each set of tasks is a checklist that will be used for correcting the reports. Review this carefully before submitting your reports. In order to get a passing grade of G on the lab report all boxes must be checked "Yes".

# Lab

## Background information

Some bacteria can exist in two different biological states. The first state is characterized by fast growth and reproduction. It is typically this state that we observe when we grow bacteria in scientific labs under ideal circumstances, e.g. with lots of food and no predators/toxins. Interestingly, when we grow bacteria in the lab, we often notice that a subset of bacteria are in a second biological state. In the second state, the bacteria grow very slowly, if at all. In addition, if we start with only bacteria in the slow-growing state and give them food, we find that there will eventually be some in the fast-growing state. This makes one wonder why such slow-growth states exist—is it an unfortunate accident or is there any benefit to it?

We observe that when a toxin like an antibiotic is applied to the bacteria, those in the fast-growing state seem to die more frequently than those in the slow-growing state. This curious observation leads you to think that the slow-growth state might be beneficial to bacteria in that it allows them to survive harsh environments, like those with antibiotics. You propose the hypothesis below.

> **Hypothesis:** *Bacteria that switch between the two biological states have an advantage over those that exist in only one state (either always fast-growing or always slow-growing).*

To evaluate the validity of your hypothesis, you build a simple mathematical model that applies your knowledge of linear algebra. You write down a simple set of finite difference equations that describe how many bacteria there will be in each biological state after some fixed period of growth in a good environment (say 30 minutes with lots of food and no antibiotics/toxins).

$$
\begin{aligned}
F_{t+1} &= (2-p)F_t + .25pS_t \\
S_{t+1} &= pF_t + (1.25 - .25p)S_t
\end{aligned}
\tag{1}
$$

Here, the $F_{t+1}$ and $S_{t+1}$ variables represent how many bacteria there are in the fast-growing state and the slow-growing state, respectively, after a fixed period of growth (from $t$ to $t+1$, where $t$ is an integer that corresponds to time). The equations show that the number in each state depend on the number of bacteria that you started with: $F_t$ and $S_t$. In addition, you define a variable $p$ that represents the *p*robability that a bacteria will switch states, i.e. switch from fast-growing to slow-growing or vice versa. Since $p$ is a probability it can take any value between 0 and 1.

With this simple system, you have separated your endeavor into four parts each with a series of tasks that you must complete. Proceed in order as later tasks rely on completion of prior tasks.

---

## Part 1: Understand the basic model

You have built the mathematical model shown in Equation set (1). Now you must demonstrate that you understand the model and that it behaves in a way that you expect.

(1a) The two equations from Equation set (1) can be formulated as a product of matrix-vector multiplication: $v_{t+1} = Av_t$ where $A$ is a matrix, $v_t$ is the vector $[F_t; S_t]$, and $v_{t+1}$ is the vector $[F_{t+1}; S_{t+1}]$. Write down the appropriate matrix $A$ when $p = 0.1$.

(1b) Write MATLAB code that performs the matrix-vector multiplication when $v_t = [1\ 0]$. What happened? Explain why this occurred in terms of linear algebra. (This is an extremely common occurrence in MATLAB)

(1c) Write MATLAB code that performs the matrix-vector multiplication when $v_t = [1; 0]$. What is $v_{t+1}$ in this case? Submit your MATLAB code.

(1d) You need a type of experimental control to show that your model behaves well. You realize that if you start with no organisms in either state then you should not end up with organisms after a period of growth. Show this. Describe what you did and submit your MATLAB code.

(1e) You want another type of experimental control to show that your model behaves well with regards to $p$. You realize that $p = 0$ is somehow special. Describe why this is (i.e. what happens when $p = 0$) and how you can use it to show that your model behaves well.

(1f) Use the resulting $v_{t+1}$ from (1c) to perform the matrix-vector multiplication again. In other words, compute $v_{t+2} = Av_{t+1}$ when starting originally from $v_t = [1; 0]$. Explain how this compares with `A*A*v` or `A^2*v`? Note: `v` is the same as $v_t$. In sample MATLAB code we drop the subscripts.

(1g) Multiplication by $A$ describes what happens after a fixed period of growth. Thus, the result of $A * A * \ldots * A * v_t$ is a vector whose components are the number of bacteria in the fast-growing and slow-growing states after many periods of growth. The total number of bacteria is simply the sum of the components of the vector, which in MATLAB can be computed by `sum(v)`. Use a `for` loop to have MATLAB compute the final number of bacteria after $k = 1, 2, \ldots 25$ rounds of growth, where $k$ corresponds to $v_{t+k}$ or equivalently $\underbrace{A * A * A * \ldots * A}_{\text{k times}}$.

Plot the total number of bacteria versus the number of growth periods. Submit the plot, your MATLAB code to generate the data and plot, and a description of why the overall trend of the graph makes sense in terms of an $A$ that represents a favorable growth environment with lots of food.

(1h) Let $k = 25$. For $k = 25$ successive growth periods, consider different values of $p$. The variable $p$ is a probability so $0 \le p \le 1$. Which value of $p$ gives the most bacteria? Provide an explanation for what you did to arrive at your conclusion and give the supporting data which could either be a well-labelled plot, MATLAB code, or a mathematical proof. Hint: There are many acceptable ways to do this. One brute force way involves using a `for` loop to iterate $p$ between 0 and 1. If you do this then remember that MATLAB only accepts integer indices so using both "`for p=0:.01:1`" and "`data(p)`" will cause an error. One workaround involves using something like "`pvalues=[0:.01:1]`" with "`for i1=1:length(pvalues)`" and "`p=pvalues(i1)`". Also note that $A$ changes depending on what $p$ is.

---

## Task checklist for Part 1

| Tasks | Yes | No | Comments |
|---|---|---|---|
| *(1a) Written matrix A* | | | |
| *(1b) Description of what happened and why* | | | |
| *(1c) The correct value of* $v_{t+1}$ | | | |
| *(1c) MATLAB code* | | | |
| *(1d) Description of how you performed the control* | | | |
| *(1d) MATLAB code* | | | |
| *(1e) Description of why* $p = 0$ *is special and how it can be used as a control* | | | |
| *(1f) Correct value of* $v_{t+2}$ | | | |
| *(1f) Explanation of comparison to other commands* | | | |
| *(1g) Plot with labels* | | | |
| *(1g) MATLAB code that generated the data and produced the plot* | | | |
| *(1g) Description of results and why they match intuition* | | | |
| *(1h) Correct value of p* | | | |
| *(1h) Description of procedure and supporting data* | | | |

## Part 2: Foray into eigenvalue analysis

(2a) We call a vector normalized if its magnitude is equal to 1. So a vector $b$ of length $n$ is normalized if $\sum_i^n (b_i^2) = 1$. To normalize a vector whose magnitude is not 1, simply rescale it. One easy way of doing it in MATLAB is to divide a vector $b$ by `norm(b)`. The result of `b/norm(b)` is a normalized vector. Normalize $v_{t+1}$ where $v_{t+1}$ is the product of $Av_t$ when $v_t = [1; 0]$ and $p = 0.1$. Hint: You can easily check that it is normalized using MATLAB.

(2b) Use the `eig` command to find the eigenvalues and eigenvectors of the matrix $A$. What is the largest eigenvalue and what is its corresponding eigenvector?

(2c) Call the largest eigenvalue of A, $\lambda_{\max}$. For $k = 1, 2, \ldots 25$, compute $(\lambda_{\max})^k$. Plot the results on a graph where $k$ is on the horizontal axis. Compare this plot with the one from (1g). Are they very similar? Are they very different?

(2d) Let $u_{\max}$ represent the eigenvector corresponding to the largest eigenvalue of $A$. We want to measure what happens to an initial vector $v_t = [1; 0]$ after being multiplied by $A$ many times. To do this we will measure the distance between vectors by computing the Euclidean distance (using the `norm` command in MATLAB). So to compute the distance between $v_{t+1}$ and $u_{\max}$, simply use `norm(v-u)`, where `v` is $v_{t+1}$ and `u` is $u_{\max}$. Compute the distance between $u_{\max}$ and the vector $v_{t+k}$ that results from $k = 1, 2, \ldots 25$ periods of growth, i.e. $\underbrace{A * A * \ldots * A}_{\text{k times}} v_t$. Plot the distance versus $k$ and submit your MATLAB code.

(2e) When taken together, what do the results from (2c) and (2d) indicate happens during prolonged growth in the same environment $A$? In other words, what is the role of the largest eigenvalue and its associated eigenvector?

## Task checklist for Part 2

| Tasks | Yes | No | Comments |
|---|---|---|---|
| *(2a) Normalized* $v_{t+1}$ | | | |
| *(2b) Largest eigenvalue* | | | |
| *(2b) Eigenvector for the largest eigenvalue* | | | |
| *(2c) Plot of* $(\lambda_{\max})^k$ | | | |
| *(2c) Comparison of plots* | | | |
| *(2d) Plot of the distance between vectors* | | | |
| *(2d) MATLAB code* | | | |
| *(2e) Explanation of the results from 2c and 2d* | | | |

## Part 3: Evaluate the hypothesis

Let us assume there are a different set of equations when the environment has a toxin or antibiotic in it.

$$\begin{aligned} F_{t+1} &= (0.6 - 0.1p)F_t + 0.1pS_t \\ S_{t+1} &= 0.1pF_t + (1.1 - 0.1p)S_t \end{aligned} \tag{2}$$

Using this system, do the following:

(3a) The two equations from Equation set (2) can be formulated as a product of matrix-vector multiplication: $v_{t+1} = Bv_t$ where $B$ is a matrix, $v_t$ is the vector $[F_t; S_t]$, and $v_{t+1}$ is the vector $[F_{t+1}; S_{t+1}]$. Write down the appropriate matrix $B$ when $p = 0.1$.

(3b) Write MATLAB code that performs the matrix-vector multiplication with $B$ when $v_t = [1; 0]$. What is $v_{t+1}$ in this case?

(3c) If there is one period of growth in $B$, which $p$ gives the most bacteria $(F_{t+1} + S_{t+1})$? Hint: does $p$ matter? Justify your response.

(3d) If the antibiotic/toxin is present for 2 periods of growth so the environment is $B*B$, consider different values of $p$ between 0 and 1. Which value of $p$ produces the most bacteria? Submit your MATLAB code that shows how you arrived at the answer.

(3e) Suppose for (3d), you found $p = 0$ gives the most bacteria, describe what this suggests for your hypothesis.

(3f) Suppose for (3d), you found $p > 0$ gives the most bacteria, describe what this suggests for your hypothesis.

(3g) You have a doubting inner voice about your finding in (3d). You wonder if the initial condition matters. Redo the analysis except use $v_t = [0; 1]$. What is the $p$ that produces the most bacteria? There is an important implication from this finding. Describe what it means in terms of whether you have confirmed your hypothesis. Recall that the hypothesis specifies a benefit to being able to produce both biological states ($F$ and $S$).

---

## Task checklist for Part 3

| Tasks | Yes | No | Comments |
|---|---|---|---|
| *(3a) Written matrix B* | | | |
| *(3b) The correct value of* $v_{t+1}$ | | | |
| *(3c) Answer to the effect of p in one round of growth and justification.* | | | |
| *(3d) Correct value of p* | | | |
| *(3d) MATLAB code* | | | |
| *(3e) Written implications of* $p = 0$ *from 3d* | | | |
| *(3f) Written implications of* $p > 0$ *from 3d* | | | |
| *(3g) Correct value of p when initial conditions are different* | | | |
| *(3g) Written implications of results.* | | | |

---

## Part 4: Insight

You have a flash of insight and wonder what if the environments fluctuate between $A$ and $B$ so that sometimes the antibiotic is present and sometimes it is absent?

(4a) Use $(B*B*B*A*A*A)^{10}$ as the matrix for the growth environment. This corresponds to fluctuating 10 times between three periods of growth in $A$ and then three periods of growth in $B$. What is the $p$ that produces the most bacteria when $v_t = [1; 0]$? Submit your MATLAB code.

(4b) Repeat the analysis from (4a) but use the starting condition $v_t = [0; 1]$. What is the $p$ that produces the most bacteria?

(4c) What do your results from (4a) and (4b) say about your hypothesis?

(4d) You now wonder what happens if you increase the duration in each environment, i.e. how long the organisms grow in the $A$ and $B$ environments. Use $(B*B*B*B*B*B*A*A*A*A*A*A)^{10}$ as the matrix for the growth environment. This corresponds to fluctuating 10 times between six periods of growth in $A$ and then six periods of growth in $B$. What is the $p$ that produces the most bacteria when $v_t = [1; 0]$? Submit your MATLAB code.

(4e) Repeat the analysis from (4d) but use the starting condition $v_t = [0; 1]$. What is the $p$ that produces the most bacteria?

(4f) What happened as you increased the duration of growth in the environment to the $p$ that produces the most bacteria? Hint: You should see a trend if you do the same type of analysis with the matrix: $(B*B*B*B*B*B*B*B*B*A*A*A*A*A*A*A*A*A)^{10}$.

(4g) Based on your results state a new hypothesis about the relationship between the $p$ that produces the most bacteria and the duration in environments $A$ and $B$, i.e. "As the duration in environments $A$ and $B$ increases…". Note that this new hypothesis could be supported by your analyses in Part 1 and Part 2.

---

## Task checklist for Part 4

| Tasks | Yes | No | Comments |
|---|---|---|---|
| *(4a) Correct p* | | | |
| *(4a) MATLAB code* | | | |
| *(4b) Correct p* | | | |
| *(4c) Description of how these results relate to the hypothesis* | | | |
| *(4d) Correct p* | | | |
| *(4d) MATLAB code* | | | |
| *(4e) Correct p* | | | |
| *(4f) Written description of results* | | | |
| *(4g) Written new hypothesis* | | | |