

May 17-18, 2021 – Hands-On Intro to the Basics of Scientific Programming
Prof. Joshua Weitz, Prof. James C. Gumbart, and the QBioS Cohort

Would you like to be able to build a computational model of a living system but don't know how? Then you are in the right spot.

This tutorial is meant to be interactive... that is you should be reading, typing, thinking, and asking questions. The materials are adapted from a semester long course entitled "Foundations of Quantitative Biosciences" developed by Prof. Joshua Weitz in Fall 2016/2017/2018/2019/2020 as the cornerstone class for the QBioS Ph.D. at Georgia Tech. The materials have been adapted to focus on epidemics modeling and to account for the greater variety of backgrounds of students in the workshop.

Today we will focus on the basics of coding that can help you build models... whether of gene expression, cellular dynamics, game theory, or some other problem linked to dynamics of living systems. Let's get started!

1 Getting Started

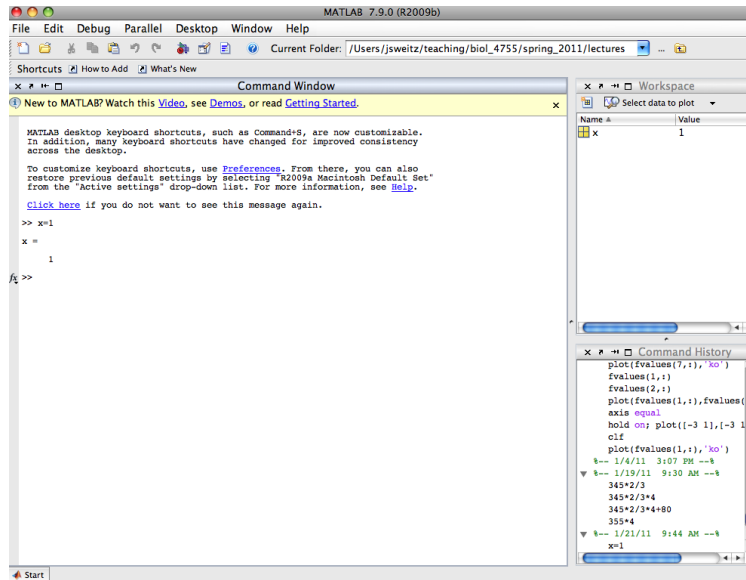
MATLAB is both a numerical computing environment and a programming language. We will be using MATLAB in this laboratory guide to simulate and model the dynamics of living systems from molecules, cells, organisms, to ecosystems. MATLAB is particularly good at:

- Matrix manipulation
- Numerical methods
- Algorithms
- Graphical output

and is increasingly good at:

- Symbolic computation
- Interfacing with other languages
- String manipulation
- Speed of `for` loops (as compared to C)

It can also be used for interactive work using the MATLAB Live Editor (similar to the iPython notebook). So, let's get started! This chapter will help you gain practical experience using MATLAB. When you open MATLAB, you should see a set of windows that looks like this:



The key elements are the Command Window, Workspace and Command History. The command window is where you enter commands, and you should see a prompt that looks like this:

```
>>
```

You can do basic math at this prompt, for example

```
>> 3+4
```

```
ans =
```

```
7
```

and

```
>> exp(1)
```

```
ans =
```

```
2.7183
```

MATLAB has **many** functions that are built in which can be used like a calculator. For example, to learn more about the exponential function:

```
>> help exp
```

```
EXP Exponential.
```

```
EXP(X) is the exponential of the elements of X, e to the X.
```

```
For complex Z=X+i*Y, EXP(Z) = EXP(X)*(COS(Y)+i*SIN(Y)).
```

```
See also expm1, log, log10, expm, expint.
```

```
Overloaded methods:
```

```
lti/exp
```

```
codistributed/exp
```

```
fints/exp
```

```
Reference page in Help browser
```

```
doc exp
```

Many functions have names that you expect (how do you think you should calculate cosine or sine of a value, for example)? Try it out!

If you don't know the name of the function you can use the command **lookfor** or **help**.

MATLAB is not just a calculator. It is also a programming language that can store values in memory. For example, the command

```
>> x=3
x =
    3
```

tells MATLAB that the variable x has the value 3 and now every time you use “x”, MATLAB will substitute the value 3, for example:

```
>> y=x+1
y =
    4
```

It is very important to realize the “=” sign does not mean that MATLAB checks to see if the two sides are equal to each other. Instead, MATLAB interprets the = sign to assign the value on the right to the variable on the left. If you want to check the truth of a particular statement – is x equal to 3, or alternatively, is y equal to 3 – then you would type:

```
>> x==3
ans =
    1
>> y==3
ans =
    0
```

The double “==” sign tells MATLAB to logically compare what is on the left with that on the right and return 1 if true and 0 if false. Note that if you don’t want MATLAB to report back the answer every time, you can use the semicolon

```
>> y=x+1;
```

MATLAB can also handle arrays of values (for example a vector or a matrix). The simplest way is to use the colon, which defines a sequential vector, for example

```
>> v=1:5
v =
    1     2     3     4     5
```

you can also modify the increments by adding a step in the middle

```
>> w=1:2:9
w =
    1     3     5     7     9
```

Any entry can be examined using the parentheses

```
>> w(3)
ans =
    5
```

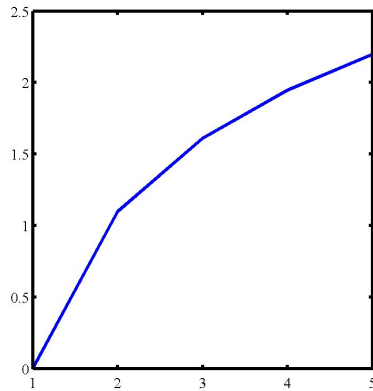
and basic math can be performed automatically on vectors (and matrices), for example

```
>> log(2*w)
ans =
    0.6931    1.7918    2.3026    2.6391    2.8904
```

MATLAB can also plot graphs, surfaces, and more. To create a simple plot, use the plot command

```
>> plot(v,log(w))
```

which leads to:



2 Building “Programs” from “Scripts” and “Functions”

Once you have a lot of commands, it will get exhausting typing them again and again (especially when you make mistakes). Instead, you will want to use a “script”. A script is a list of commands in a file that you can execute directly from the command window. To create a script go to the File menu and select New > M-file. Now type in a few commands, such as:

```
% my_first_file.m
% Create some vectors
x = 1:0.1:10;
y1 = exp(0.5*x);
y2 = exp(0.6*x);

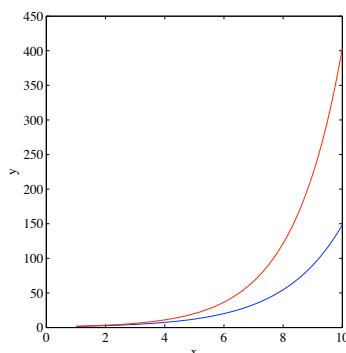
% Plot the vectors
figure          % Generate a new figure
hold on        % Allow multiple plots on the same figure
plot(x,y1,'k'); % Use a black line
plot(x,y2,'r'); % Use a red line
xlabel('x'), ylabel('y'); %Label the axes

% Save the image to a file
print -dpdf my_first.pdf
```

Save this file as my_first_file.m in the same folder you’re currently working in. Now go to the MATLAB prompt and type

```
>> my_first_file
```

and it should execute the commands in the script to yield the figure:



The problem with this script is that changing the arguments in the exponential functions requires editing the script and then re-running the code. It would be more convenient to designate a variable change from the command window and have the code automatically update its output. The problem is that a script cannot return a variable or accept a variable as input. To do so requires a “function”. Functions are program files that can be called from the Command window, can accept inputs, and return outputs. To start one, open a new M-file and type:

```
function dNdt=logGrowth(t,N)
% function dNdt=logGrowth(t,N)
%
% logGrowth gives the growth rate of a population of size N at time t
% Usage:dNdt=logGrowth(t,N)
r =0.5;
K =100;
dNdt =r*N.*(1-N/K);
```

Now the new function can be accessed just like one of MATLAB’s built-in functions, for example,

```
>> N = 0:110;
>> figure
>> plot(N, logGrowth(0,N),'k-');
>> xlabel('N'),ylabel('dN/dt');
```

This gives an upside down parabola, denoting that growth rate is positive between 0 and 100 and negative when N is greater than 100. Note that the argument `t` is not used in the `logGrowth` function. Not all inputs have to be used. We will update this function later to both accept and utilize all inputs.

3 Getting Started with Core Techniques

3.1 Create loops

There are many loops that enable the repetition of a fixed set of commands. The two central loops examined here are “for” and “while” loops. Both loops start with a keyword such as `for` or `while` and they end with the keyword `end`. The `for` loop allows us to repeat certain commands many times with a “counter” variable. Here is one example:

```
>> for j=1:4 % counting
    j+2 % the statement you want to repeat
end
```

A counter variable can be incremented by any real number, e.g.,:

```
>> for j=1:-0.1:0 % counting
    1-j % the statement you want to repeat
end
```

For loops can also span an arbitrary set of values:

```
>> for j=[1 3 5 7]
    j-1 % the statement you want to repeat
end
```

Challenge Problem: Exercise on Matrices

Define a random matrix A of size 3-by-3. Use a double for loop to calculate the square of the entries in A and store the values in another matrix B. (Hint: type “help rand” in MATLAB if you don’t know how to define a random matrix)

A “while” loop is the most robust of loops; formally speaking it is a universal loop such that all loops can be built using it. That’s a topic for some other class – for our purposes, both while and for loops may be appropriate depending on the circumstance. The `while` loop repeats a sequence of commands as long as some condition is met. For example, given a number n , the following code will return the smallest non-negative integer a such that $2^a \geq n$.

```
function a = smallexp(n)

a = 0;
while 2^a < n
    a = a + 1; % statement to execute if the condition is met
end
```

such that this command can be invoked from the MATLAB interface:

```
>> a = smallexp(4)
```

```
a =
    2
```

Note that in the above example we used the conditional statement, $2^a < n$ to decide whether the statement within the while loop should be repeated. Such conditional statements are also used in “if” statements. The relational operator used here is “<”, which means “less than”. Other relational operators that are available in MATLAB include:

```
> greater than
<= less than or equal
>= greater than or equal
== equal
~= not equal
```

Simple conditional statements can be combined by logical operators

```
(&&, ||, ~)
```

into compound expressions such as the following:

```
(y > 1) && (x == 6)
```

3.2 Make decisions

Now, let’s suppose you want your code to make a decision. In many circumstances, a series of `if` statements will suffice. The general form in MATLAB is as follows:

```
if expression1
    statements1
elseif expression2
    statements2
else
    statements3
end
```

Challenge Problem: Recursive Factorial

Finish the following pseudo-code that gives the factorial of a positive number n , using the recursive formula $n! = n(n-1)\dots 1$

```
function N = factorial_recur(n)

if #
    N = 1;
else
    N = #;
end
```

3.3 Go fast, i.e., “vectorize”

Remember: **for** loops are SLOW in MATLAB. One way to make your MATLAB run faster is to **vectorize** the algorithm you use in the code. Vectorization can be done by converting **for** and **while** loops to equivalent vector or matrix operations. A simple example would be the following (**tic** and **toc** are used as a stop watch):

```
x = 1;
tic
for k = 1:1001
    y(k) = log10(x);
    x = x + .01;
end
toc

tic
x = 1:.01:11;
y = log10(x);
toc
```

You should find that the second command set is faster, even if it returns exactly the same output. However, for more complicated code, vectorization is not always so obvious. Some of the most commonly used functions for vectorizing can be found in the Help browser of MATLAB.

3.4 Find values you want to know about

Given an array X , the command

```
>> find(X)
```

returns the indices of all nonzero elements of X . You can also use a logical expression to define X . For example,

```
>> find(X>2)
```

returns the indices corresponding to the entries of X that are greater than 2. Notice that X could also be a matrix. In this case, the function returns the linear indices as if the matrix was stored as a single column of elements (try it!). The ‘elementwise’ logical operators (& and |) can be used to locate the entries that satisfy more than one logical expressions. These commands will be important in solving the next challenge problem.

Challenge Problem: Finding Elements in Matrices

Build a 5-by-5 random matrix A using the command

```
>> A = rand(5)
```

Find the indices of entries whose values are smaller than $1/4$ and bigger than $1/6$. Check the answer with your (virtual) neighbors or study partners.

3.5 Save and load data

The keywords in MATLAB are straightforward: **save** and **load**. By typing

```
>> save filename
```

MATLAB will save all variables in the current workspace in 'filename' (which you should replace with an actual filename). If you do not specify an extension to the filename, MATLAB uses .mat. When you want to load all the variables from the file specified by filename, just type

```
>> load filename
```

Please keep in mind that these file formats are binary and proprietary, i.e., not human-readable. If you want to save a particular file format use the help to save variables in standard, comma-separated value file format. There are also alternative ways to load and print data.

4 Numerically Integrating Differential Equations

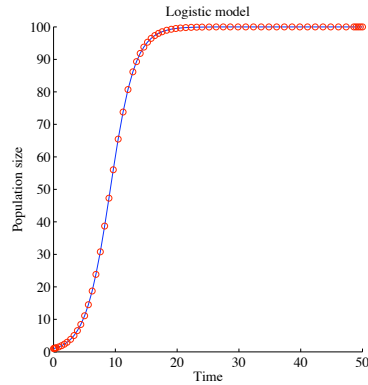
In this class we can use MATLAB to numerically solve differential equations, like the logistic growth equation, even when such solutions are not available analytically. The most-used MATLAB program which does the integration is called `ode45`. Here is a script that integrates the `logGrowth` function. We will return to this multiple times in the course.

```
% Numerical solution of the logistic equation
t0 =0; % Initial time
tf =50; % Final time
N0 =1; % Initial population size
[T, Nint] = ode45(@logGrowth, [t0 tf], N0); %Numerically integrate
% Actual solution
r =0.5;
K =100;
Nact =(N0*exp(r*T))./(1+N0*(exp(r*T)-1)/K); %Actual solution
% Plot results
figure;
hold on
plot(T,Nint); % Plot numerically integrated solution
plot(T,Nact,'ro'); %Plot actual solution
xlabel('Time'), ylabel('Population size'),title('Logistic model');
print -dpdf logPlot.pdf
```

Save this script as `intLog.m` and run it

```
>> intLog
```

which yields the plot



Here are some important points to keep in mind:

- MATLAB solves ordinary differential equations of the form

$$\frac{d\vec{y}}{dt} = \vec{f}(\vec{y})$$

where $\vec{f}(\vec{y})$ is a vector of functions. The default solver used in these computational laboratories is **ode45**.

$$[TOUT,YOUT] = ODE45(ODEFUN,TSPAN,Y0)$$

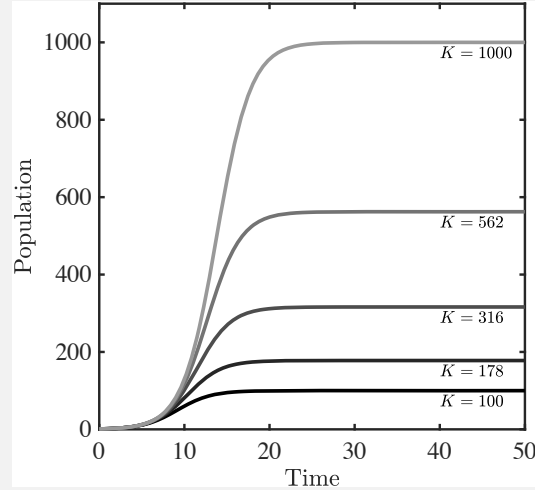
where

1. TOUT \rightarrow time of output
2. YOUT \rightarrow value of variable
3. ODEFUN \rightarrow name of function containing the dynamics
4. TSPAN \rightarrow time limits for integration
5. Y0 \rightarrow initial conditions

and refer to **help ode45** for more information. The key idea is that MATLAB has a built-in function that numerically integrates a differential equation that must be specified – by you. It integrates the system of equations over a range of time given initial conditions. Let's give this a try, albeit by going farther: accepting additional parameters as input to the dynamical system.

Challenge Problem: Variables and Differential Equations

Read the `ode45` documentation and modify your logistic growth model to accept the parameters r and K . Setting $r = 0.5$, vary the value of K over 1 order of magnitude and compare the results using a plot command. If your code works, the resulting figure should look something like this:



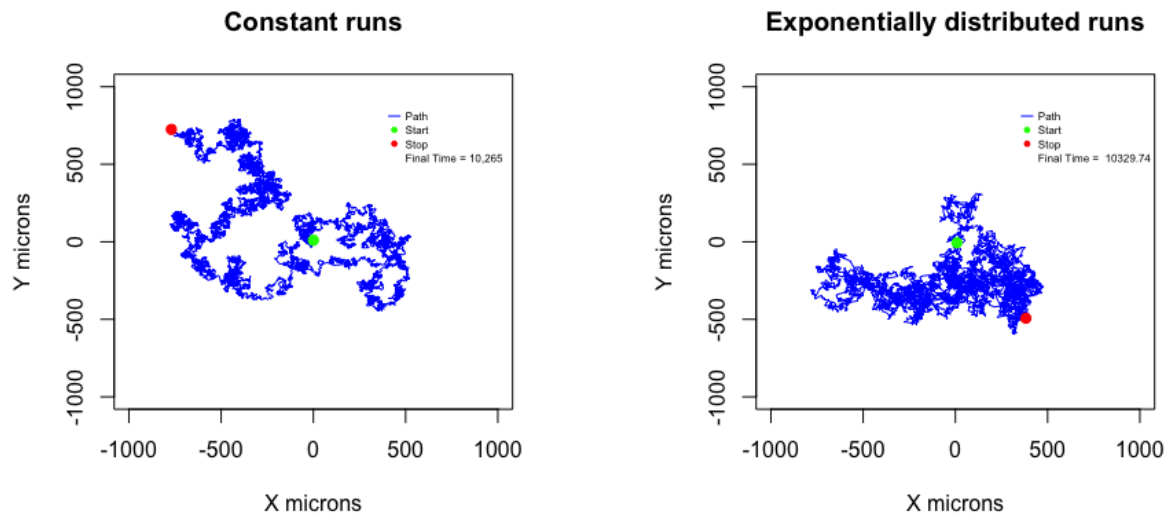
5 Advanced - Individual-Based Simulations

As a final challenge for those more experienced in MATLAB, try and develop a simulation of “diffusion” in which a bacteria swims at a speed of $10 \mu\text{m/s}$ with each “run” lasting 1 second. In this case the direction of the next run is random. How long will it take, on average, for the bacteria to travel 1mm away from its source? As you may recall, the average time for a diffusing particle or organism to travel a squared distance x^2 on average should be:

$$T = \frac{x^2}{D} \quad (1)$$

It is true that any particular bacteria may travel much farther. What do we mean by *average* here? Indeed, the average distance traveled by bacteria is 0. That is the bacteria is just as likely to go to in any particular direction. But if one squares the distance from the origin, you will find that the average squared distance increases linearly with time.

If you develop code, can you visualize the results? If you change the length in equation, how does T change? Can you confirm the scaling and estimate D ? What happens if the durations of each run is exponentially distributed, so that runs last on average 1 second, but can vary in duration? Did the answer with respect to travel time to reach 1mm change in a quantitative or qualitative way? We will get much further into these ideas later in the class. Here are examples of two runs, one with constant duration and one with exponentially distributed durations:



6 Take-home messages

- Building computational models of living systems requires a willingness to code, tinker, try, and try again.
- Expanding your computational toolset will take time.
- Mistakes are part of the learning process and indeed, will help reinforce common pitfalls.
- With time, your objective should be able to code for the most part without having to open a help forum or to try and search for the answer in your favorite search engine.
- The basic skills presented here are not comprehensive, instead they are a gateway to further exploration.
- The modules that follow build on a few of the methods in this laboratory, with a focus on direct integration of computational modeling as an integral part of the quantitative biosciences approach to science.
- You can do it! Really.