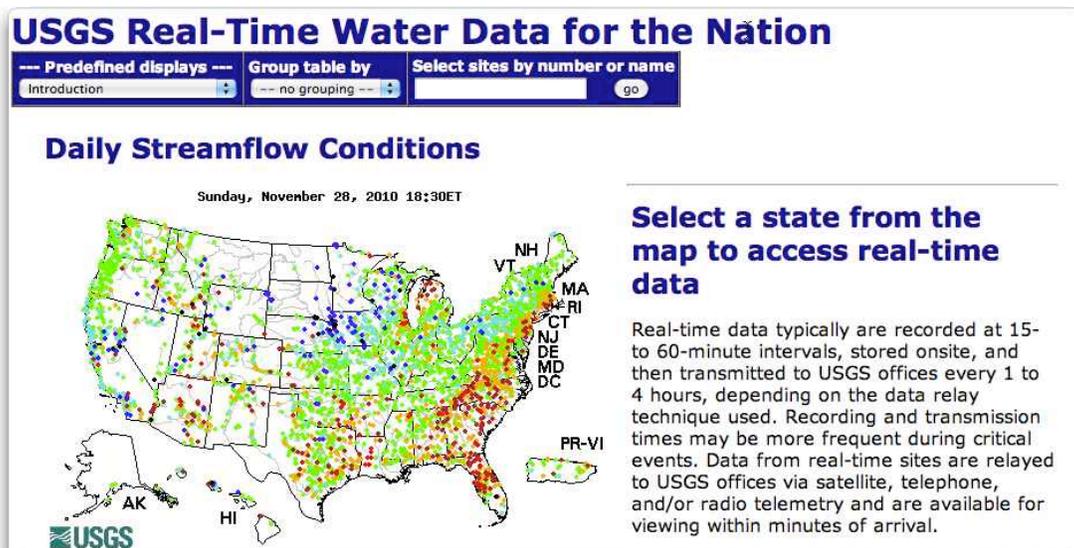

Final Project: Part 1

Predicting Stream Height from Stream Volume Flow

Technical Problem Solving - Fall, 2010

Lansing Community College

Louise Paquette



Introduction to the Engineering Problem

In any community that has a large body of water, and in particular any community with a river, there is a continuing chance of a flood. The amount of property damage alone caused by floods can be staggering. The picture at the right is a typical flood.¹ In 1993, the national per capita cost for floods approached \$70. That



is - the cost of 1993 floods was almost \$70 for every man, woman, and child in the United States. The total cost - over \$16B... that is 16 BILLION dollars.

Managing streams and rivers in the United States is the responsibility of a number of federal and state agencies, including at a national level the US Army Corps of Engineers. A fundamental part of the task of the Corps and other agencies is analysis and prediction of the height of streams and rivers.

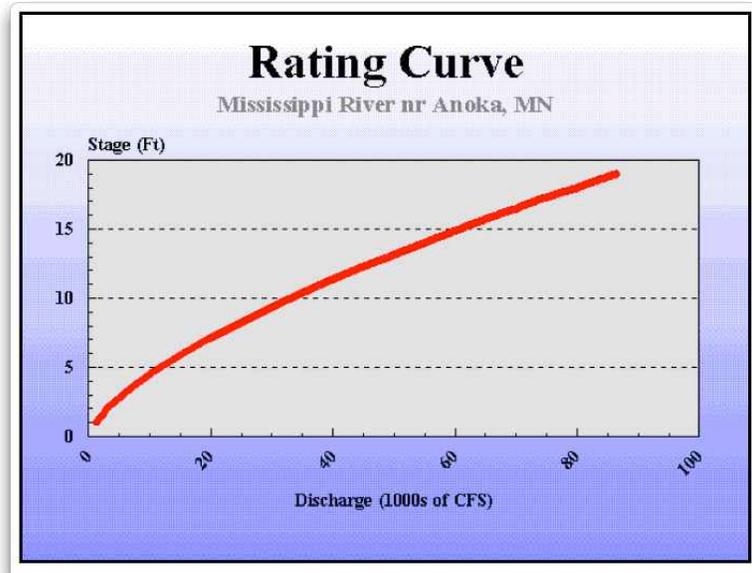
If you have a data model of the height of a stream given an amount of water moving through the stream, you can determine if the river is going to overflow its banks. And knowing that, if the stream will overflow you can invoke emergency measures such as building sandbag levies to hold back the water. The process starts with an engineer - usually a CIVIL ENGINEER - determining how high the river will get given an amount of water expected (from e.g., heavy rains upstream).²

Such a data model is conceptually simple. It simply means keeping very accurate records of the the two variables you are interested in: (a) the total flow rate of the

¹ From the National Institute for Disaster Management, <http://www.nidm.gov.in/images/flood6.jpg>.

² Determining the amount of water that will flow through a river at a given point is itself a difficult problem. It depends on where the rain fell upstream, the soil properties upstream (e.g. sand or clay), the soil condition (e.g., frozen, dry, ...) and a number of other factors. We are going to assume that this problem has been solved for us, so our problem becomes simply predict the river height given an expected river flow in our target city.

river at your location and (b) the river height at your location. Data like this is called a time series because *at the same time* you measure both (a) and (b). You make many measurements of the two variables. Then you make a plot of the river height (usually called the *stage*) versus the river flow rate



(usually called the *discharge*). An example is to the above.³ You keep careful records over a long time, then simply plot the result. The plotted relationship is a data model - a model useful for prediction that is based on empirical data.

Description of Your Term Project Part 1

Once you have the data in hand to create a rating curve, the problem is basically solved. You know how to make relatively sophisticated two-dimensional graphs using MATLAB. The heavy lifting for your project is to find suitable data for creating a rating curve for our location in Lansing, Michigan, and getting that data into MATLAB. The real heart of the project is to represent data you find on the web in a digital form within MATLAB.

Step 1: Identify data on the web for use in a rating curve

Go to <http://waterdata.usgs.gov/>. This is the main web site for the water data source of the United State Geological Survey - one of the most important federal agencies that systematically measures stream data and

³ From http://www.nws.noaa.gov/os/hod/SHManual/SHMan040_rating.htm

archives it. Since the web became a medium for data distribution, USGS has made their data publicly available.

Click on REAL TIME DATA. This will bring up a map of the United States showing all stream and river collection points of USGS. (A copy is shown on the cover page of this document.)

Now click on Michigan in the map. That will bring up a page that shows Michigan only and show the USGS recording stations more clearly. Each of the dots on the map is an interactive point in the map. Find the green point that represents the GRAND RIVER AT LANSING.⁴ The identifying number for this USGS recording site is

USGS
04113000

Make sure
you have
the page
open with
that
identifier.

Available Parameters	Output format	Days	
<input type="checkbox"/> All 2 Available Parameters for this site	<input type="radio"/> Graph	120	GO
<input checked="" type="checkbox"/> 00060 Discharge	<input type="radio"/> Graph w/ stats	(1-120)	
<input checked="" type="checkbox"/> 00065 Gage height	<input type="radio"/> Graph w/o stats		
	<input type="radio"/> Table		
	<input checked="" type="radio"/> Tab-separated		

Now on the page you have open, find the box that looks like the one to the right. This box controls the data that you will be able to see. Make sure you have the selections made in the box that are indicated above. These settings will bring up a page that shows the discharge and the gage height for the Grand River at Lansing for the last 120 days when you click GO.

Go ahead and click GO.

You will see a number of lines in the web page that start with a # symbol. Following these, is a table given. The first two columns of the data show the data source (USGS) and the recording location identifier (04113000). The third column is the date of the observations in the row; the fourth column holds the time of the observation. The fifth column indicates the time zone of the observation. The next two columns are what we are most interested in: the sixth column contains the measured height at the specified date/time, and the seventh column contains the measured river discharge at the specified date/time.

⁴ You can magnify the web page by click on <ctrl>+. That will help you locate the Grand River reporting point.

So at this point - we have the data needed - at least for a rating curve for the last 120 days. *Now we need to get it into MATLAB.*

Step 2: Get the data into a TXT file

This is an easy step.

First look at your data. The real data you are interested in starts with the first row that has USGS in it. In your browser, select ALL the data rows (from the first row starting with USGS all the way to the last element on the page).

Now open TextEdit if you are on a MAC or a plain text editor if you are on a PC. Paste the data from your clipboard into the open text editor window. Be sure you are using a font like COURIER or COURIER NEW. Doing that will help you to see the orderliness of the data.

Here is a sample set of rows from the data set like they might appear in your TXT file.

```
USGS 04113000    2010-08-26 05:45 EST    2.54    283
USGS 04113000    2010-08-26 06:00 EST    2.54    283
USGS 04113000    2010-08-26 06:15 EST    2.54    283
USGS 04113000    2010-08-26 06:30 EST    2.54    283
USGS 04113000    2010-08-26 06:45 EST    2.54    283
USGS 04113000    2010-08-26 07:00 EST    2.54    283
USGS 04113000    2010-08-26 07:15 EST    2.54    283
USGS 04113000    2010-08-26 07:30 EST    2.54    283
USGS 04113000    2010-08-26 07:45 EST    2.54    283
USGS 04113000    2010-08-26 08:00 EST    2.54    283
```

Create a new folder on your computer to keep your MATLAB program for your project in - name the folder MyFinalProject. Save your TXT file with all the data into this folder. Name your TXT file MyRawProjectData.txt.

Step 3: Start MATLAB

Once you start MATLAB, make MyFinalProject be the connected directory.

Step 4: Move the data into MATLAB

This is where things are going to get more interesting.

There are many ways to get data into MATLAB. Many times you will have data in an EXCEL file that you want to load into MATLAB. For that MATLAB gives you a specialized function

`xlsread`

Open EXCEL and enter new data into a blank workbook that is shown below:

1	2
2	16
3	54
4	128
5	250
6	432
7	686
8	1024
9	1458

Save the workbook as TrialXLS.xls (save it in 97-2004 format).

Now in MATLAB, execute the following command:

```
M = xlsread('TrialXLS.xls');
```

Simple as that - the data in your EXCEL spreadsheet is in MATLAB now. (By the way - what is the relationship between data in the first column and data in the second column?)

Look up `xlsread` in MATLAB HELP for more details.

In many cases, `xlsread` will get us through. However, for the data in MyRawProjectData.txt we are going to use a different method. The data that is in the USGS collection is very mixed. Several of the columns are numbers, but some of the other columns contain dates, times, or character data only.⁵

⁵ You have used character data (strings) to label plots before in this term.

When the data you want to get into MATLAB from EXCEL is simple, there is even an easier way to transfer it into MATLAB. Open the EXCEL file TrialXLS.xls. Select the cells with data. And COPY to your clipboard. Now in the workspace window of MATLAB, right click and select NEW. Name the newly created variable MyEXCELTest. Then double click on it in the workspace window. That will open the variable editor window of MATLAB. Right click your cursor in cell 1,1. Select PASTE EXCEL DATA. You should see the data you copied to the clipboard pasted directly in your MATLAB array, MyEXCELTest.

Keep these methods of getting numerical data into MATLAB from EXCEL in mind.

In our case - again - we cannot use the simple method.

Our data is in a tab delimited TXT file.

Open a MATLAB editor and create a function as defined below. Save the new function.

```
function M = readDataInFromTXT(fileName)
% reads data from a tab delimited TXT
% file into a MATLAB cell array.

% INPUT: fileName - the name of the TXT file to be read in
%           (assumes its in the connected directory)

% OUTPUT: M is a cell array with seven elements, one for
%           each column of data in fileName.

    fid = fopen(fileName); % open the file for reading

    theFormat = '%s %s %s %s %s %f %f';
    M = textscan(fid, theFormat);

    fclose(fid);           % close the file
```

In your MATLAB command window, invoke this function:

```
>> M = readDataInFromTXT('MyRawProjectData.txt');
```

... then just type and execute M in your command line:

```
>> M
```

You will find that M is an array with seven items in it. Each one of the items is a “cell vector” containing over 11,000 elements. Each of these “cell vectors” contains data from one of the seven columns in the original TXT file.

But this is a ***type of data*** that you have not seen before - a “cell array.”

Step 5: Stretch your mind and learn about non-numeric data in MATLAB

We have all the data from the recording station of the USGS for the Grand River in Lansing that we need now. But before we can proceed to process that data, you need to stretch yourself and learn some new things about MATLAB. In particular, you need to learn the rudiments for strings (character data) and for cell arrays in MATLAB.

You will find a slide set at the following URL that describes what you need to know both for strings and for cell arrays:

<http://breeze.msu.edu/p43585803/>

This is a PDF file, and you can download it if you want to. The slides cover (a) strings and (b) cell arrays. For your final project this term, you need to read and understand only (a) and (b) - strings and cell arrays.

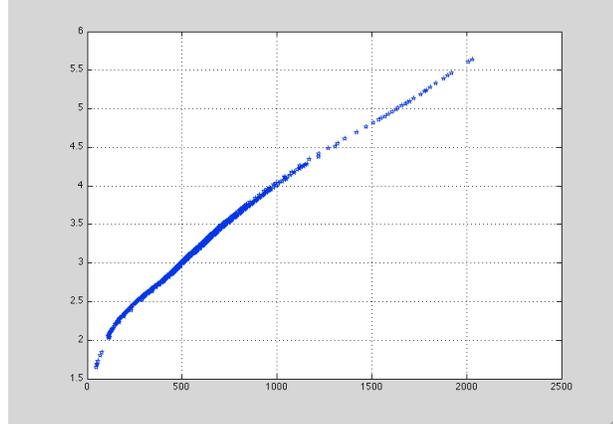
The best way for you to handle this new MATLAB capability quickly is for you to read the slides, and every place there is MATLAB code, actually execute the code in your MATLAB. This will help you become more confident of yourself on these new part of MATLAB.

Step 6: Now create your rating curve!

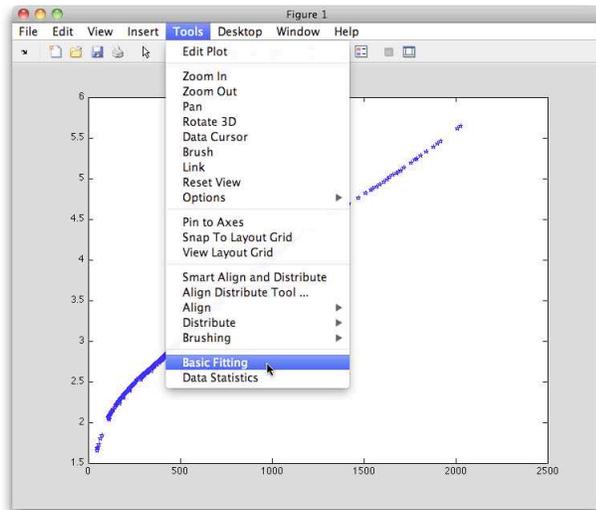
Yes - you read that right. You have all you need now to create the rating curve. Your rating curve should look similar to the one below - expect that yours of course should be properly labeled. Do not “connect the dots” in your graph. Show each data point and an unconnected dot (or some shape of your choosing).

Step 7: Use MATLAB to Create a mathematical function to Model the Rating Curve

Now that you have the rating curve shown graphically, you can go one step further, and use MATLAB to easily give you a formula that will approximate the data. You have done least squares fitting of data previously. Under the hood, the MATLAB graphing tool does the same operation.

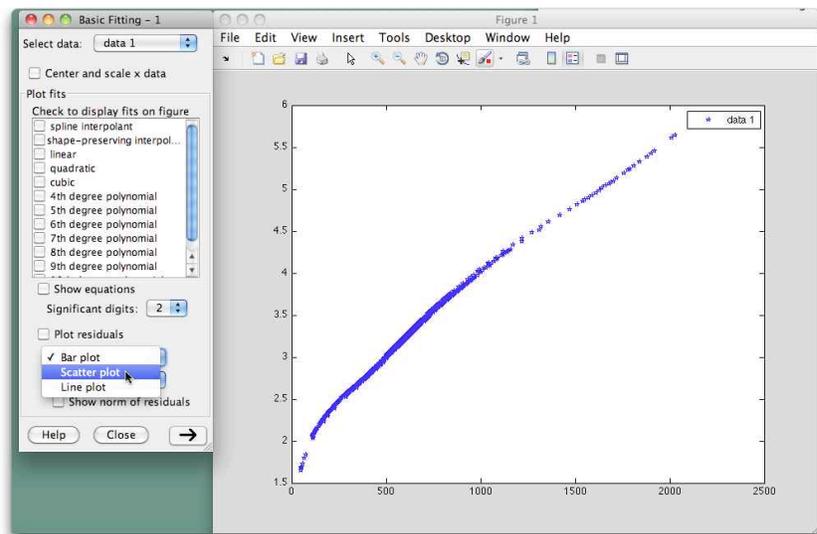


To bring up the tools for fitting a curve, pull down the tools menu from the plotting window select BASIC FITTING as shown in the diagram to the right.

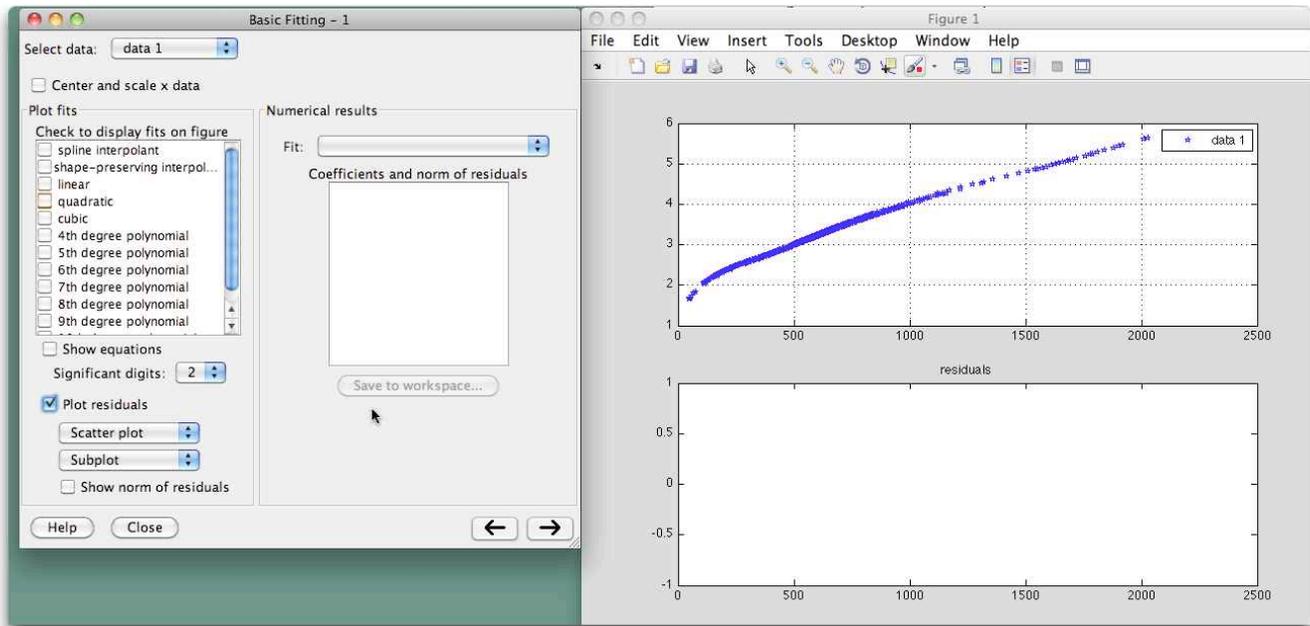


A new window will come up that will let you choose the type of curve fit you want to model with. Choosing an appropriate type of curve fit is very difficult and in fact is the topic of a whole area of

mathematics. For most engineering purposes we can simplify the process. To start, choose the type of residual plot we want to be a scatter plot. The residual is a measure for every point in the original data of

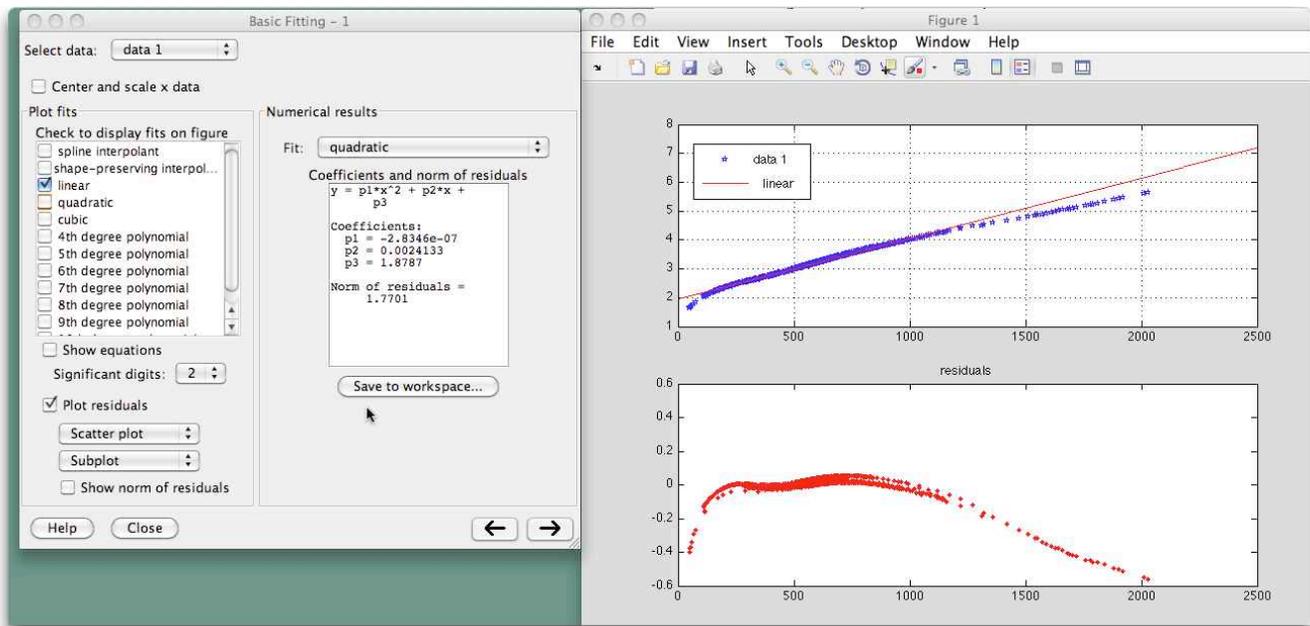


how far away the actual data is to the curve fitted function. When you click on “Plot Residuals” you will get a subplot window showing under your main window. When you plot the residuals you hope to see a more or less random pattern. It is usually not purely random. But the more random it is, the more it indicates that you have a good curve fit.



Ok - choose your first curve fit type to be LINEAR. That produces the red curve as shown, and the residuals as shown. Now in the extra pane opened there is a value for the “Norm of the residuals.” This is the single most important number to keep track of as you try higher order polynomial fits. The general idea is to choose a fit such that going to the next higher polynomial does not produce very much change in the norm of the residuals. You do not want to “over fit” the data. (Look this term up and understand it.)

Notice that you also get from MATLAB the polynomial coefficients that best fit the data for the type of curve fit you are currently trying. These coefficients give you a mathematical model (based on data) for the rating curve. You can use this model to now make predictions for the height of the Grand River given the volume flow of water.



Step 8: Limitations of the Rating Curve

In general, *interpolation* is viewed as much more on solid ground than *extrapolation*. (Be sure you understand both these terms in the context of curve fitting!)

Why would extrapolation to high volume of flow be particularly shaky with respect to the rating curve? Base your answer on a physical argument not a general mathematical argument.

Next Step

Yep - there is a next step - actually several of them. This has been PART 1. But before you go on to PART 2, you should know what the point of this project has been up to this point.

Doing a rating curve for a river is intrinsically an interest problem - and an often solved problem for engineers in one branch of Civil Engineering. So this has been a real problem that engineers solve routinely in their work.

Yet the rating curve itself is intrinsically easy. You just need the two time series of data: one time series for the height of the river and a second time series for the total flow of the river. Then the plot is easy.

The hard part is actually getting the data. Civil engineers get that data by making very systematic observations of river height and flow over many time periods. Others can grab the data from these observations from the internet in this day and age. But grabbing the data is just the start.

You saw that the data had to be put into a proper digital form in MATLAB to enable your use of the MATLAB tools for plotting and function finding given the data. That is a recurring theme in engineering - the devil is in the details, and in this case, the devil is in the process of representing the data properly in MATLAB. Of course, there is first the problem of just finding the data on the web that you might be looking for. More than anything, that is a matter of persistence and getting advice when you need it.

The specific “learning objectives” for Part 1 of your project are:

- 1) Demonstrate an ability to import data from EXCEL into MATLAB.
(Remember the side exercise to use xlsread and secondly to copy/paste data from EXCEL into MATLAB?)
- 2) Demonstrate a base level ability to represent data with cell arrays and string data in MATLAB,
- 3) Demonstrate an ability to follow a complex set of instructions to reach a technical goal.

#3 you need to put in context. No one class can help you learn all there is to know about MATLAB. But one class can help build in you the confidence that you can master new topics in MATLAB if you develop habits of “life long learning.” That is just a fancy way of saying that you learn that MATLAB is very functional as technical computing tool, and that you are willing to use MATLAB Help and find out how to use some capability in it that you need, but don’t current know how to use.

In PART 2, you will get a chance to demonstrate even more that you meet learning outcome #3 above.