CogSci 109 Fall 2007 Assignment 3 : Basic data manipulation, visualization, and code optimization (100 pts, 5 bonus)

C. Alex Simpkins

October 15, 2007

1 Description

Read this entire document before beginning the assignment.

During this assignment you will load data into matlab, and discover important contents of that data using one of the techniques we have discussed thus far. Additionally you will process more than one data set, where you will be asked to apply different techniques as appropriate (ie low pass filter, high pass filter, etc).

The assignment is broken down into steps. When a step will be used as a criterion for giving you points, the number of points and what the points correspond to is explicitly stated.

2 Formatting details (10 points total for this section)

- (1 point) Each plot should be printed such that information is redundantly encoded IF more than one plot is in the same window (ie multiple colors AND linetypes) but clear (good contrast, proper scaling of axes). Give a title to each

plot with a plot number as part of the title. When referring to the plot in the written sections, use that number.

- (1 point) Each page of the homework should be numbered
- (1 point) There should be a title page which includes the homework title, your name, the date, the course number, and the instructor
- (1 point) When commenting on a data set or plot, title each comment with the plot number, or somehow clarify which plot the comment is associated with.
- (5 points) Include at the end of your homework a printout of your matlab scripts. The code should be well commented [1] and include a header [2].
- (1 point) The homework should be neat and carefully stapled. The paper should be clean. If you use lined paper for the handwritten portions, do not turn it in with the portion torn out of a binder. Cut or otherwise remove that from the paper first.

3 Part I: Loading and displaying data sets

3.1 Download the data file

Download and unzip the folder of files (**homework3data.zip**) for this assignment at the web site in the assignments section for CogSci 109 (**http://maelabs.ucsd.edu/alex/ pages/cogsci109_07**). You should have a folder which is called homework3data, containing 5 matlab .mat files.

3.2 Load the data (10 points)

Use either the *import wizard* or the *load* command to load the data sets into a variable of your choice, such as *data*, one variable per data file. Or alternatively, you could just do each plot problem one at a time - load, display, etc. Either way, include the code to load a single set of data (or if you use another method such

[1] remember, use a percent sign to start a comment line

[2] recall that a header is a section of commented code at the beginning of your code document which describes the code and includes relevant info like author, homework number, date, class, etc

as double-clicking the file, use one or two sentences to explain how you loaded the data). This will determine the 10 points

3.3 Display the data (25 points)

Display the raw data for each of the files by making a plot of each set of raw data[3]. Each plot should have appropriate axes, titles, and legends. If one or more of the data sets is a 2x2 or larger matrix, plot the data using either the *pcolor* or *surf* commands.

Notes for each dataset:

- dataset1 (1 figure, 5pts)- plot a versus t (*plot(t,a)*), t refers to time, you don't need to plot them separately
- dataset2 (1 figure, 5pts)- no special comments
- dataset3 (1 figure, 5pts)- no special comments
- dataset4a and b (4 figures, 10 pts total)- these two datasets are related. See the comments below.

3.3.1 hw3_dataset4a and b specifics and additional questions

For the large data set (hw3_dataset4a.mat and hw3_dataset4b.mat), consider that you are consulting for a small group of doctors. They have sent you this data file with very little documentation, and asked you to make sense of it. For these two sets (**hw3_dataset4a.mat and hw3_dataset4b.mat**), try plotting one *pcolor* plot for each pair of dimensions (ie, if it is a 100x100x100 matrix, plot the first two dimensions using something like ***pcolor(squeeze(B(:, :, 1)))***) to discover what is the most salient pair to use for data visualization. Choose one or two slices of this matrix to display, and comment on what it might be.

Now load both hw3_dataset4a.mat and hw3_dataset4b.mat, and you have two variables : A and B. Add A to B (A+B), then perform a pcolor plot and comment on a possible link between the two files (why is it more appropriate to interpret the A matrix in terms of the combined plot of A and B?).

[3]We refer to a set of data as *raw* when it has not been altered in any way since recording

Show two different color maps for hw3_dataset4b.mat (using built-in matlab colormaps) as *subplots*, one of which causes additional details to be more clear. Explain this difference in terms of human perception.

3.4 Expose information by creating a custom colormap (10 points)

Create a custom color map which is not built into matlab, and is designed to highlight a specific aspect of the information in the image for the larger matrix data set (ie if it were a human hand scan, it might highlight the section of the image which shows the bones). Type *help colormap* to get more info on creating your own color map in matlab.

3.5 Comment on each data set - interpret results (2 points per set, 5 sets, 10 points total)

Briefly comment on the characteristics of each data set. What can you immediately recognize from the data (amplitude, noise levels, math functions, matrix characteristics - if a matrix, or structural entities)?

4 Part II: Filtering (35 points, 5 bonus possible)

4.1 Filter the data (5 pts)

Perform, using the functions presented in class, and starting with the code handouts on the web site, a low pass filter and high pass filter on the hw3_dataset1.mat data set. You should experiment with a few window sizes for the low pass filter to find one which removes the high frequency components, and correspondingly a high pass filter which removes the low frequency components effectively.

4.2 Explain the filters (5 pts)

Write down and explain using a few sentences what the filtering functions do. How might you improve the speed of execution of these functions, if possible?

4.3 Respond to questions (10 pts)

Use the *tic* and *toc* matlab commands to measure the computational time to perform each filter, and comment on the differences you see. Why does the time it takes to compute the filter change as you increase the window size for the moving average filter? Does the value of the time change for the recursive low pass filter as you vary the value of a? Explain why this changes or does not change.

4.4 Interpret results of filtering (10 pts)

After filtering each signal, what can you say about the filtered data? Comment on items such as amplitude, basic function types you can see in the data, etc.

4.5 Get your bonus points here!

(BONUS: The moving average filter as presented in class is not the most efficient it can be, especially as applied to matlab coding. Rewrite the moving average function to be more efficient. Show that you improved the computational efficiency by using the tic and toc commands to measure the execution time of your new function and compare it to the old one).

4.6 Filter Plots (5 pts)

Plot the filtered results in the same figure window and plot as the raw hw3_dataset1.mat data using a different color and linetype which is clear (perceptually).

END OF HOMEWORK 3