CISC 271, Winter 2014
Assignment #2
Due on Thursday, April 3, 2013

There are two questions for you to complete in this assignment. Read the following details and instructions carefully before you proceed to work on the problems.

Marking system:

- The value of each problem appears in parenthesis besides the problem title.
- The quality of your `Matlab` code will be considered. Your code should be appropriately indented, sufficiently commented, and otherwise be appropriate software.

What to turn in:

- You will email your answers, to the Teaching Assistant (TA), as a zip file.
- The zip file will contain a cover text file (`cover.txt`) and as many `Matlab` files (`.m`) as needed. Your entire report will be in a single PDF file (`xxxxx.pdf`) , where the x's stand for your Queen's NetId. The text file must include your name, your Queen's id , and your answers. You may include notes, comments on the problems, and assumptions you have made in the PDF file.
- The zip file will be named `a2_xxxxx.zip`, where the x's stand for your Queen's id.

Policies:

- You must complete these questions individually.
- Although you are allowed to discuss the questions with other students, you must write your own answers and `Matlab` code.
- The code **must** run on the CasLab computers; if it does not run you may receive zero marks for the assignment.
- Lateness policy applies starting the next calendar day of the submission deadline (as specified on Moodle), at a rate of 20% off the assignment value per calendar day.

1. Problem 1: (8 points) Principal Components Analysis of biomedical data

Note: This problem uses data from human subjects that was kindly provided by Professor Kevin Delusio from the Department of Mechanical and Materials Engineering at Queen's University. You are not expected to read his journal article or know the details of the data.

The data are in two files, `a2q1group1.dat` and `a2q1group2.dat` that can be read using the Matlab `load` command. Each file is an array of 30×101 numbers, where each row is a signal as described in the course notes. The data have been normalized and shows how each subject's knee flexes and extends during a cycle of normal walking.

Please note that the course notes, and the code within it, treat the signals as columns. if you use the instructor's code then you will have to transpose the data. If you use other code, for example if you implement PCA yourself, you will need to carefully decide whether the code treats an individual's signal as a row vector or as a column vector.

The signals are knee motion angles, collected using a calibrated computer vision system as each of the subjects walked in a laboratory setting. One group was 30 individuals with arthritis; the other group was age-matched individuals who seemed to be in good musculoskeletal health and thus acted as a control group.

The problem to be solved is: which group is which, and how can you justify your choice? To solve this you must use PCA. You may use the code provided that the instructor wrote, or Matlab functions, or any other code that you wish. If you do not use the instructor's code then you must carefully document where the code was from and how the code worked.

To solve this problem for each group, use PCA to find the mean signal and the eigenvalues/eigenvectors of the covariance matrix. Next, determine how many principal components are needed to "capture" most of the variation from the mean signal (the number is different for each group). Finally, reduce each individual signal to a sum of principal components and determine how the error between the reconstruction and the original signal.

A general guideline for PCA is to capture 90% of the spectrum. These are exceptionally good signals, with 7 eigenvalues covering 98% of the differences, so you should use very few principal components and cover about 90% (*but no more than 93%*) of the variation.

It is recommended that you plot the mean signals, principal components, and reconstruction errors. You do *not* need to plot all 60 signals; a brief summary, such as the average norm of reconstruction errors, is sufficient.

If you found any particular signal was poorly reconstructed, please bring it to our attention in your report. The data are provided in the `ZIP` file for the assignment. The instructor's Matlab code is in the notes for the classes during which PCA was discussed.

Problem 2: Solving Least-Squares Geometrical Problems (7 points)

In image-guided surgery, geometrical problems often arise that can be solved in the least-squares sense. This question is about two such problems: finding the focal point of an X-ray imaging system, and finding the tip of a sharp surgical probe. Solutions were presented during class and in the notes, so this question is about their implementation.

2(a): Crossing Lines in Space (5 points)

In this problem, you will be given a set of data that are on the XY plane and a matching set of data in space. The spatial data also happen to lie on a plane, which is tilted and offset from the XY plane. The two data sets represent a common configuration: the output of image-processing where the spatial data have been sensed as the planar data using an X-ray imaging device. For simplicity, we will model the X-rays as coming from a point source.

The data given to you are "excellent" in the sense that they produce lines that almost perfectly cross (within calculation errors). Your problem is threefold:

(i) Apply small random deviations to the planar data,

(ii) Estimate the focus as the point that minimizes line error, and

(iii) Statistically summarize your results.

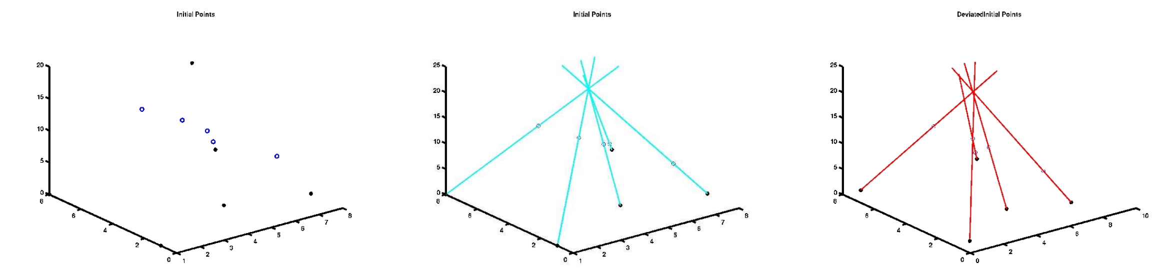
The random deviations can be from any zero-mean distribution you wish; the instructor used a uniform distribution in formulating this assignment but you just need to clearly state your methods. The raw data, the resulting lines, and deviated data with the nearly crossing lines are given in Figure 1.

In the `ZIP` file for the assignment, there is a Matlab workspace containing the data. if you execute the Matlab command

```
load a2q2data
```

then you will find the variables `PlanarPoints` and `SpatialPoints` that are the data.

In your report, *clearly* and concisely summarize your findings. You may analyze the errors as the distances from the least-squares focal point to the lines, or the skew-line distance, or any other reasonable error measure.



(A) **(B)** **(C)**

Figure 1: Data to fit to a set of lines. (A) The basic data, stars on the X-Y plane and circles in space. (B) The data fit lines that exactly cross. (b) Randomly deviated data fit lines that cross imperfectly or not at all.

2(b): Hemispherical Data (2 points)

A common problem in setting up an image-guided surgical system is calibrating the surgical instruments. In a wide variety of surgical procedures, a sharp "probe" instrument is used. This assignment question is how to calibrate such an instrument.

The instruments typically are located in space by a tracking device; the technology is immaterial here but is most often uses optoelectronic or electromagnetic sensing. To calibrate the tip of a sharp probe, we can model the tip as one point and each tracked element over time as a set of spatial points.

The calibration process usually requires that the tip be fixed in space; one way is to place it in a specially manufactured tiny "divot" in rigid material. As the probe is moved, the tracked location is constrained to lie on the surface of a sphere. The problem for this question is to test a method for finding the position and radius of such a sphere. A drawing of part of a sharp probe is Figure 2 and an image of a probe used in surgeries is in Figure 3.

To solve this problem, you must generate zero-mean random data from the surface of part of a sphere. It is recommended that you sample from a hemisphere and then perform zero-mean random deviations as you did for Question 2(a). In practice, the radius of the sphere is between 100mm and 150mm with the data centered about 500mm from the tracking device. The tracking devices that the instructor uses have various error magnitudes, ranging from 0.15mm to 0.75mm depending on the device and its configuration. In practice, no fewer than 10 points are collected and up to 100 may be used in certain applications. As above, the probability distribution of your errors does not matter provided that it has a zero mean.

In your report, *clearly* and concisely summarize your findings. The most common analysis is the root-mean-square (RMS) error of the fit, but you may use another measure provided that you clearly describe the measure and why you chose it.

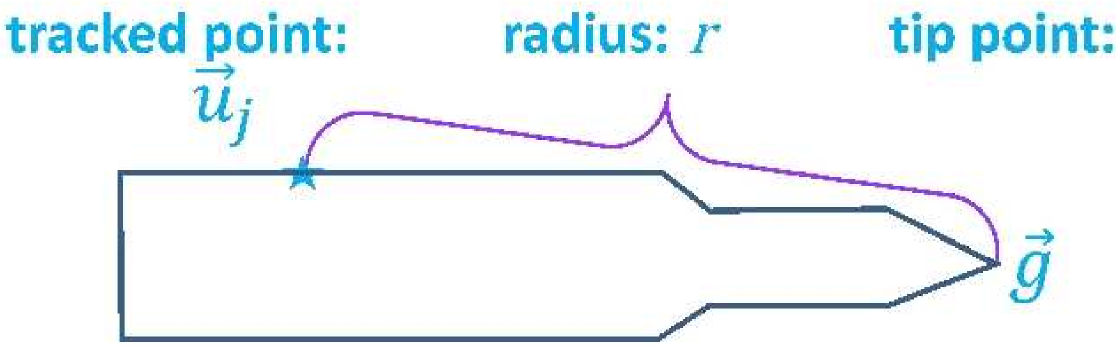


Figure 2: Drawing of a sharp probe. The tip is assumed to be stationary in tracking coordinates at the vector $\vec{g}$. The points, captured over the duration of a calibration experiment, are vectors $\vec{u}_j$ that the tracking device reliably estimates. The nominal radius r is the Euclidean distance from the tip of the probe to the tracked point.

Figure 3: An annotated photograph of a sharp probe. The tracked point was an infrared light-emitting diode; the tip was surgical stainless steel (air hardening) that was ground to sharpen it.