SIO 239 2008 Homework #5 (Second Half)

[1] Show that the projection matrix written as

$$\mathbf{I} - \mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T$$

can be alternately written as

$$\mathbf{I} - \mathbf{U}(:, 1..m)\,\mathbf{U}^T(:, 1..m)$$

where m is the number of nonzero singular values in the svd decomposition of $\mathbf{A}$. Notice that this projection operator, which projects onto the null space of $\mathbf{A}$ is obtained from very basic ingredients in the first formula and an elegant but more advanced construct in the second. From the viewpoint of using `matlab`, since the svd algorithm is readily available, there is no reason not to prefer the second for practical applications, noting that for larger matrices you should definitely use the "economy size" version of the svd call where applicable.

[2] I have prepared a data file `tide.mat`, which contains four columns and 88320 rows. The values are the level of Lake Superior (in meters) recorded at six minute intervals over the course of a year at four stations around the lake. These are (in the order of the columns): Duluth MN, Grand Marais MN, Marquette MI (Coast Guard Station), and Pt. Iroquois MI.

You will find that there are six missing data entries, dropouts for unspecified reasons in the NOAA entries. For the first part of this exercise, eliminate the six rows affected. Next remove the running mean from the series (that is, the average over the four columns) and find the svd of the result. From the singular values, it is a fair approximation that most of the signal is contained in the first two components out of four. Look at the corresponding first two columns of `v` (where I am assuming you use the command: `[u,s,v]=svd(tide,0);`). Comment on the meaning of the values you note in the first column. Note the signs and magnitudes of the entries in the second column and then locate the four tide stations on a map. Explain.

For the next part I want you to take the fft of `u(:,2)`, but the `u` constructed above is not uniformly spaced because of the data dropouts. So please next devise what you consider as a plausible way to provide substitute values for the six missing entries so that you have then a uniformly spaced series. Recompute each `u` for this series and take the Fourier transform of all four. Looking in the range from 600 to 800 (that is, the array argument), see what spectral feature of note you can identify as a significant feature of one of the svd modes. Determine the dimensional period of this feature. Returning to the map, convert this period into an equivalent velocity.

Assuming the dataset starts at midnight GMT on 26 Sept, 2007, what is the date and time of the peak event associated with `u(:,2)`? What is the amplitude of this event in meters? What, for contrast, is the standard deviation of this signal (again, that associated with the second singular value mode) as recorded at Grand Marais? (One can consult meteorological data and find a causal relation.)