QMDA Homework 6

Empirical Orthogonal Function (EOF) analysis

Read into the Matlab workspace (use the "load" function) the file "MOR_EOF_228_Z.mat" (in "Files/Homework" on CourseWorks). You will then have the following in your workspace:

- **Z**, a matrix of 228 rows by 81 columns. Each row of Z is a 81-point bathymetric profile across the axis of a mid-ocean ridge. The ridge axis is located at the 41st data point. Depths in the profiles are measured every km, with the point in the first column of Z at –40 km and the last point at +40 km from the ridge axis. A reference depth has been subtracted from each profile (row of **Z**).

The EOF analysis is described in a paper by Small (1994) and the results for this data set of 228 profiles are in Small (1998). Both papers are in "Files/Readings" on CourseWorks.

A) Calculate the singular value decomposition (SVD) of the data matrix **Z** in Matlab using the 'econ' option (which gives the most compact matrices) as follows:

```
>> [U,S,V] = svd(Z,'econ');
```

Plot the first five singular vectors (the first five columns of V) with an *x*-coordinate that corresponds to distance from the ridge axis. Compare your results to the inset with the first five singular vectors (called "Modes") in the small inset in Figure 2 of Small (1998). Do your results match? (Note that the Matlab SVD algorithm may have obtained the negative of any one of the singular vectors in in Figure 2.)

B) Reproduce the main plot in Figure 2 of Small (1998) by plotting the values on the diagonal of **S** as fractions of the sum of all the values in **S**. Make sure your results match those in Figure 2.

C) The caption of Figure 2 of Small (1998) states "*The singular values indicate the distribution of variance over the spatial modes given by the decomposition* […] *The five spatial modes corresponding to the largest singular values* […] *account for 44% of the variance in the dataset*." Given what we discussed in class, this statement is incorrect. Explain why and generate a new figure that plots the actual fraction of the total variance explained by each singular vector. Calculate and report a corrected fraction of the total variance due to the first five singular vectors. (Hint: it should be more than 44%.)

D) Define a diagonal matrix of singular values $\mathbf{S}_p$ that contains the same first five singular values of **S** but has zeros everywhere else. You can then compute an approximation to the data matrix as

```
>> Zp = U*Sp*V';
```

Compare rows 1, 80, and 120 of **Z** and $\mathbf{Z}_p$ by plotting them in the same figure as a function of an *x*-coordinate that corresponds to distance from the ridge axis. For example, plot the row of **Z** as a black line and the corresponding row of $\mathbf{Z}_p$ as a red line. Comment on the differences between the original profiles in **Z** and those computed using only the first five singular vectors in $\mathbf{Z}_p$. Include the plots in your report.