Problem Set 6: Solutions

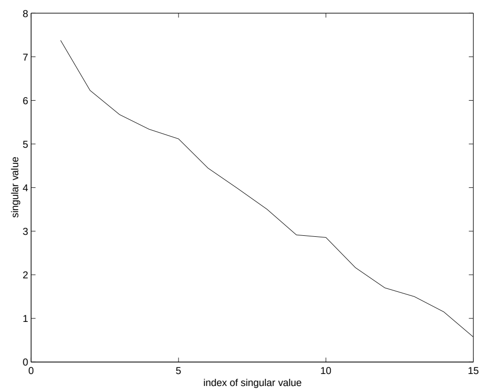
1. Demonstrate that the relative sizes of the singular values of a matrix provide a measure of the variance represented by each singular mode. To do this, first generate an N by M matrix $\mathbf{G}$ consisting of random numbers. Then compute the singular value decomposition of $\mathbf{G} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$. (Matlab will do this with the "svd" command.)

a. Plot the singular values.

b. Define a matrix $\mathbf{H}_k = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k$, which is the reconstruction of $\mathbf{G}$ using only the kth singular value and singular vectors. Compute the variance of all values in $\mathbf{H}_k$ and $\mathbf{G}$. What is the relationship between the variance of $\mathbf{H}_k$ and the kth singular value?

c. Now show (using equations) that the L2 norm of $\mathbf{H}_k$ is the kth singular value, λ_k.

Solution:
a. Singular values should decrease with increasing value.



b. Empirically, we note that $\sum_i \sum_j H_k(i,j)^2 = \lambda(k)^2$
c. To show that the result in (b) is generally true, note that $||H_k|| = ||U_k\lambda_k V_k^T|| = \lambda_k||U_k V_k^T V_k U_k^T|| = \lambda_k||U_k U_k^T||$. Since the norm of U_k is 1, this is equal to λ_k.

2. The temperature along a transect in the ocean is believed to satisfy a linear rule $\theta = ar+b$, where r is latitude measured in kilometers from a reference point, and a, b are constants. Measurements at sea yield the following table of values:

$$\begin{aligned} r &= 0, \quad \theta = 10 \\ r &= 1, \quad \theta = 9.5 \\ r &= 2, \quad \theta = 11.1 \\ r &= 3, \quad \theta = 12 \end{aligned}$$

a. Using ordinary least-squares, find an estimate of a, b, and estimate the noise in each measurement, and the standard errors of the estimated values of a, b.

b. Using the singular value decomposition (you can use an SVD program if you wish), also find an estimate of a, b, and its standard error. Which of the equations proved most important to the solution? Are the elements of the solution fully resolved?

Solution:
a. By ordinary least squares, the solution will be of the form $\mathbf{a} = (\mathbf{G^T G})^{-1}\mathbf{G^T}\theta$. Thus $a = 0.76$ and $b = 9.51$. The misfit $r = \theta - G\mathbf{a}$. Since we have no information about the size of the errors, let's assume that χ^2 should be $N-2$. Then $\sigma^2 = \sum r^2/(N-2) = 0.441$. This implies that the error bars are $\sigma_a = 0.297$ and $\sigma_b = 0.556$.

b. To solve the problem by singular value decomposition, first compute the svd(G). The solution $\mathbf{a} = \mathbf{V\Lambda^{-1}U'}\theta$ and turns out to be the same as in (a). To determine whether the solution is fully resolved, compute VV'. This is a 2 by 2 identity matrix, so the solution is fully resolved. To determine which equations matter most for the solution, compute the data resolution matrix UU'. This is a 4 by 4 matrix:

$$UU' = \begin{bmatrix} 0.7 & 0.4 & 0.1 & -0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.3 & 0.4 \\ -0.2 & 0.1 & 0.4 & 0.7 \end{bmatrix}$$

The diagonals of this matrix are largest for the first and fourth elements, meaning that these have the greatest impact on the solution.

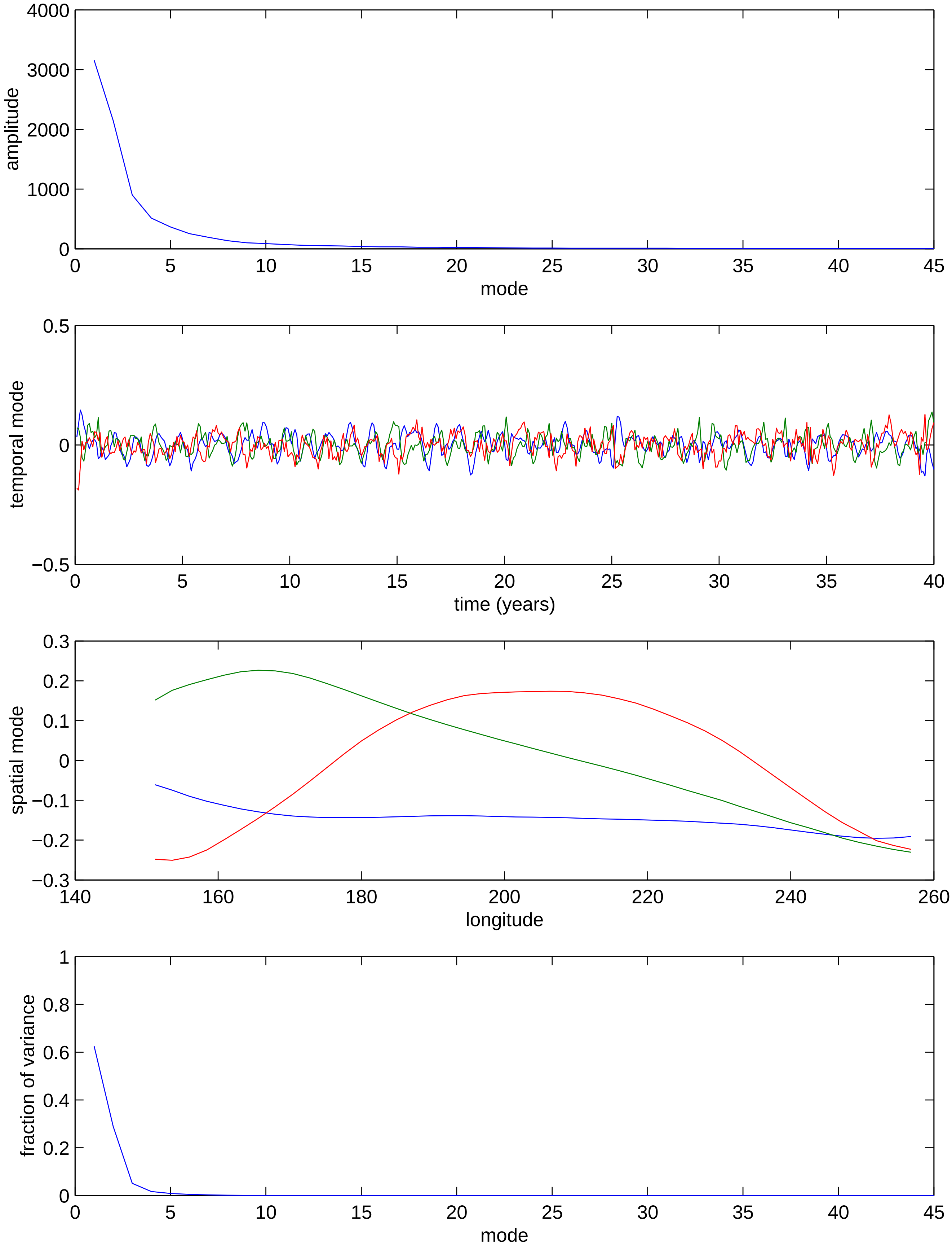
3. From the course web site:
http://www-mae.ucsd.edu/∼sgille/sio221b/homework.html
download the matlab data set "hw6.mat" or the individual ASCII files "ueq.dat" and "loneq.dat". The variable "ueq" is a 480 by 45 matrix containing monthly zonal velocities on the equator at 22 m depth from a 40 year run of the NCAR CSM Ocean model. The variable "loneq" identifies the longitudes of the velocities.

a. Represent "ueq" in empirical orthogonal modes. An easy way to do this is to use the singular value decomposition. Explain your methodology.
b. Plot the first 3 temporal modes and the first 3 spatial modes of the EOF.
c. What fraction of the variance in "ueq" is captured by the first 3 EOF modes?

Solution:
a. To compute an EOF decomposition on this regularly spaced data, compute the singular value decomposition of "ueq". (See top panel of figure for singular values.)

b. See middle two panels. First mode is blue, second green, third red.

c. The fraction of variance is $\lambda^2/\sum\lambda^2$. See bottom panel.


amplitude
mode
temporal mode
time (years)
spatial mode
longitude
fraction of variance
mode