9. EOFs, SVD

A common statistical tool in oceanography, meteorology and climate research are the so-called empirical orthogonal functions (EOFs). Anyone, in any scientific field, working with large amounts of data having covariances, is almost inevitably led to EOFs as an obvious tool for reducing the number of data one must work with, and to help in obtaining insight into the meaning of the covariances that are present. The ubiquity of the tool means, unfortunately, that it has been repeatedly reinvented in different scientific fields, and the inventors were apparently so pleased with themselves over their cleverness, they made no attempt to see if the method was already known elsewhere. The consequence has been a proliferation of names for the same thing: EOFs, principal components, proper orthogonal decomposition, singular vectors, Karhunen-Loève functions, optimals, etc. (I'm sure this list is incomplete.)

The method, and its numerous extensions, is a useful one (but like all powerful tools, potentially dangerous to the innocent user), and a brief discussion is offered here. The most general approach of which I am aware, is that based upon the so-called singular value decomposition (e.g., Wunsch, 1996 and references there). Let us suppose that we have a field which varies, e.g., in time and space. An example (often discussed) is the field of seasurface temperature (SST) in the North Pacific Ocean. We suppose that through some device (ships, satellites), someone has mapped the anomaly of SST monthly over the entire North Pacific Ocean at 1° lateral resolution for 100 years. Taking the width of the Pacific Ocean to be 120° and the latitude range to be 60° each map would have approximately $60\times120 = 7200$ gridded values, and there would be 12×100 of these from 100 years. The total volume of numbers would then be about 7200×1200 or about 9 million numbers.

A visual inspection of the maps (something which is *always* the first step in any data analysis), would show that the fields evolve only very slowly from month-to-month in an annual cycle, and in some respects, from year-to-year, and that much, but perhaps not all, of the structure occurs on a spatial scale large compared to the 1° gridding. Both these features suggest that the volume of numbers is perhaps much greater than really necessary to describe the data, and that there are elements of the spatial structure which seem to covary, but with different features varying on different time scales. A natural question then, is whether there is not a tool which could simultaneously reduce the volume of data, and inform one about which patterns dominated the changes in space and time? One might hope to make physical sense of the latter.

Because there is such a vast body of mathematics available for matrices, consider making a matrix out of this data set. One might argue that each map is already a matrix, with latitude and longitude comprising the rows and columns, but it suits our purpose better to make a single matrix out of the entire data set. Let us do this by making one large column of the matrix out of each map, in some way that is arbitrary, but convenient, e.g., by stacking the values at fixed longitudes in one long column, one under the other (we could even have a random rule for where in the column the values go, as long it is the same for each time—this would just make it hard to figure out what value was where). Then each column is the map at monthly intervals, with 1200 columns. Call this matrix $\mathbf{E}$, which is of dimension M =number of latitudes times the number of longitudes by N, the number of observation times (that is, it is probably not square).

We now postulate that any matrix $\mathbf{E}$ can be written

$$\mathbf{E} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{9.1}$$

that is as the product of three matrices. $\mathbf{\Lambda}$ is a $M \times N$ diagonal matrix (in a generalized sense for a non-square matrix). Matrix $\mathbf{U}$ is square of dimension M, and $\mathbf{V}$ is square of dimension N. $\mathbf{U}, \mathbf{V}$ have the special properties of being "orthogonal",

$$\mathbf{U}\mathbf{U}^T = \mathbf{I}_M, \mathbf{U}^T\mathbf{U} = \mathbf{I}_M, \mathbf{V}\mathbf{V}^T = \mathbf{I}_N, \mathbf{V}^T\mathbf{V} = \mathbf{I}_N \tag{9.2}$$

that is to say, in particular the columns of $\mathbf{U}$ are mutually orthonormal, as are the columns of $\mathbf{V}$ (so are the rows, but that proves less important). $\mathbf{I}_N$ is the identity matrix of dimension N, etc. The matrices $\mathbf{U}, \mathbf{V}$,$\mathbf{\Lambda}$ can be shown, with little difficulty to be determined by the following relations:

$$\mathbf{E}\mathbf{E}^T\mathbf{u}_i = \lambda_i^2\mathbf{u}_i, 1 \leq i \leq M, \ \mathbf{E}^T\mathbf{E}\mathbf{v}_i = \lambda_i^2\mathbf{v}_i, 1 \leq i \leq N. \tag{9.3}$$

That is to say, the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{E}\mathbf{E}^T$, and the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{E}^T\mathbf{E}$. They are related to each other through the relations,

$$\mathbf{E}\mathbf{v}_i = \lambda_i\mathbf{u}_i, 1 \leq i \leq N, \ \mathbf{E}\mathbf{u}_i = \lambda_i\mathbf{v}_i, 1 \leq i \leq M. \tag{9.4}$$

Note that in (9.3,9.4), M, N are in general different, and the only way these relationships can be consistent would be if all of the $\lambda_i = 0, i > \min(M, N)$ (this is the maximum number of non-zero eigenvalues; there may be fewer). By convention, the λ_i and their corresponding $\mathbf{u}_i, \mathbf{v}_i$ are ordered in decreasing value of the λ_i.

Consider $\mathbf{E}^T\mathbf{E}$ in (9.3) This new matrix is formed by taking the dot product of all of the columns of $\mathbf{E}$ with each other in sequence. That is to say, $\mathbf{E}^T\mathbf{E}$ is, up to a normalization factor of $1/M$, the covariance of each anomaly map with every other anomaly map and is thus a covariance matrix of the observations through time and the $\mathbf{v}_i$ are the eigenvectors of this covariance matrix. Alternatively, $\mathbf{E}\mathbf{E}^T$ is the dot product of each row of the maps with each other, and up to a normalization of $1/N$ is the covariance of the structure at each location in the map with that at every other point on the map; the $\mathbf{u}_i$ are thus the eigenvectors of this covariance matrix.

Consider by way of example, $\mathbf{E}^T\mathbf{E}$. This is a square, non-negative definite matrix (meaning its eigenvalues are all non-negative, a good thing, since the eigenvalues are the λ_i^2, which we might hope would be a positive number). From (9.1, 9.2),

$$\mathbf{E}^T\mathbf{E} = \mathbf{V}\mathbf{\Lambda}^2\mathbf{V}^T = \sum_{i=1}^{N} \lambda_i^2 \mathbf{v}_i \mathbf{v}_i^T, \tag{9.5}$$

Eq. (9.5) is an example of the statement that a square, symmetric matrix can be represented exactly in terms of its eigenvectors. Suppose, only $K \leq N$ of the λ_i are non-zero. Then the sum reduces to,

$$\mathbf{E}^T\mathbf{E} = \sum_{j=1}^{K} \lambda_i^2 \mathbf{v}_i \mathbf{v}_i^T = \mathbf{V}_K \mathbf{\Lambda}_K^2 \mathbf{V}_K^T, \tag{9.6}$$

where $\mathbf{\Lambda}_K$ is truncated to its first K rows and columns (is now square) and $\mathbf{V}_K$ contains only the first k columns of $\mathbf{V}$. Now suppose that some of the λ_i are very small compared, e.g., to the others. Let there be K' of them, much larger than the others. The question then arises as to whether the further truncated expression,

$$\mathbf{E}^T\mathbf{E} \sim \sum_{i=1}^{K'} \lambda_i^2 \mathbf{v}_i \mathbf{v}_i^T = \mathbf{V}_{K'} \mathbf{\Lambda}_{K'}^2 \mathbf{V}_{K'}^T, \tag{9.7}$$

is not still a good approximation to $\mathbf{E}^T\mathbf{E}$? Here, $\mathbf{V}_{K'}$ consists only of its first K' columns. The assumption/conclusion that the truncated expansion (9.7) is a good representation of the covariance matrix $\mathbf{E}^T\mathbf{E}$, with $K' << K$ is the basis of the EOF idea. Conceivably K' is as small as 1 or 2, even when there may be hundreds or thousands of vectors $\mathbf{v}_i$ required for an exact result. An exactly parallel discussion applies to the covariance matrix $\mathbf{E}\mathbf{E}^T$ in terms of the $\mathbf{u}_i$.

There are several ways to understand and exploit this type of result. Let us go back to (9.1). Assuming that there are K non-zero λ_i, it can be confirmed (by just writing it out) that

$$\mathbf{E} = \mathbf{U}_K \mathbf{\Lambda}_K \mathbf{V}_K^T = \sum_{i=1}^{K} \lambda_i \mathbf{u}_i \mathbf{v}_i^T \tag{9.8}$$

exactly. This result says that an arbitrary $M \times N$ matrix $\mathbf{E}$ can be represented exactly by at most K pairs of orthogonal vectors, where $K \leq \min(M, N)$. Suppose further, that some of the λ_i are very small compared to the others Then one might suppose that a good approximation to $\mathbf{E}$ is

$$\mathbf{E} \sim \mathbf{U}_K \mathbf{\Lambda}_{K'} \mathbf{V}_{K'}^T = \sum_{i=1}^{K'} \lambda_i \mathbf{u}_i \mathbf{v}_i^T. \tag{9.9}$$

If this is a good approximation, and $K' << K$, and because $\mathbf{E}$ are the actual data, it is possible that only a very small number of orthogonal vectors is required to reproduce all of the significant structure in the data. Furthermore, the covariances of the data are given by simple expressions such as (9.7) in terms of these same vectors.

The factorizations (9.1) or the alternative (9.8) are known as the "singular value decomposition". The λ_i are the "singular values", and the pairs $(\mathbf{u}_i, \mathbf{v}_i)$ are the singular vectors. Commonly, the $\mathbf{v}_i$ are identified as the EOFs, but they can equally well be identified as the $\mathbf{u}_i$; the choice is arbitrary, depending only upon how one seeks to interpret the data.

Eq. (9.9) can be discussed slightly differently. Suppose that one has an arbitrary $\mathbf{E}$. Then if one seeks to represent it in L pairs of orthonormal vectors $(\mathbf{q}_i, \mathbf{r}_i)$

$$\mathbf{E} \approx \sum_{i=1}^{L} \alpha_i \mathbf{q}_i \mathbf{r}_i^T, \tag{9.10}$$

then the so-called Eckart-Young-Mirsky theorem (see references in Wunsch, 1996) states that the best choice (in the sense of making the norm of the difference between the left and right-hand sides as small as possible), is for the$(\mathbf{q}_i, \mathbf{r}_i)$ to be the first L singular vectors, and $\alpha_i = \lambda_i$.

Exercise. Interpret the Karhunen-Loève expansion and singular spectrum analysis in the light of the SVD.

Exercise. (a) Consider a travelling wave $y(r,t) = \sin(kr + \sigma t)$, which is observed at a zonal set of positions, $r_j = (j-1)\Delta r$ at times $t_p = (p-1)\Delta t$. Choose, $k, \sigma, \Delta r, \Delta t$ so that the frequency and wavenumber are resolved by the time/space sampling. Using approximately 20 observational positions and enough observation times to obtain several temporal periods, apply the SVD/EOF analysis to the resulting observations. Discuss the singular vectors which emerge. Confirm that the SVD at rank 2 perfectly reproduces all of the data. The following, e.g., would do (in MATLAB)

```
» x=[0:30]';t=[0:256]';
  » [xx,tt]=meshgrid(x,t);
  » sigma=2*pi/16;k=2*pi/10;
  » Y=sin(k*xx+sigma*tt);
  » contourf(Y);colorbar;
```

(b) Now suppose two waves are present: $y(r,t) = \sin(kr + \sigma t) + \sin((k/2)\,r + (\sigma/2)\,t)$. What are the EOFs now? Can you deduce the presence of the two waves and their frequencies/wavenumbers? (c) Repeat the above analysis except take the observation positions r_j to be irregularly spaced. What happens to the EOFs? (d) What happens if you add a white noise to the observations?

Remark 3. *The very large literature on and the use of EOFs shows the great value of this form of representation. But clearly many of the practitioners of this form of analysis make the often implicit assumption that the various EOFs/singular vectors necessarily correspond to some form of normal mode or simple physical pattern of change. There is usually no basis for this assumption, although one can be lucky. Note in particular, that the double orthogonality (in space and time) of the resulting singular vectors may necessarily require the lumping together of real normal modes, which are present, in various* *linear combinations required to enforce the orthogonality. The general failure of EOFs to correspond to physically interpretable motions is well known in statistics (see, e.g., Jolliffe, 1986). A simple example of the failure of the method to identify physical modes is given in Wunsch (1997, Appendix).*

Many extensions and variations of this method are available, including e.g., the introduction of phase shifted values (Hilbert transforms) with complex arithmetic, to display more clearly the separation between standing and travelling modes, and various linear combinations of modes. Some of these are described e.g., by von Storch and Zwiers (1999). Statistics books should be consulted for the determination of the appropriate rank and a discussion of the uncertainty of the results.