Feb 18: Empirical Orthogonal Functions

These notes loosely based on Emery and Thompson, Section 4.3.

The terminology Principal Components (PC) and Empirical Orthogonal Functions (EOF) are generally used interchangeably in the earth sciences. Oceanographers almost always refer to EOFs, whereas meteorologists and climatologists mix both terms.

Eigenvector theory

In its simplest formulation, EOFs are eigenvectors of the data covariance matrix. The eigenvectors are commonly referred to as modes or loadings.

For the sake of illustrating EOFs here we consider a data set described by a space and a time coordinate, though there are many examples of EOF analysis with different coordinate frameworks (such as two spatial dimensions).

Suppose the data are arranged in a matrix with ‘time’ being column dimension and ‘space’ being the row dimension. Then covariance can be computed over time as is typically done for time series data. It can be shown that the eigenvectors of the covariance matrix are related to the Singular Value Decomposition of the data matrix. The eigenvalues and singular values are the same.

The Singular Value Decomposition of the data matrix effectively by-passes computation of the covariance matrix which can be large if the spatial dimension of the data set is dense. The choice is an issue of computational convenience. Issues are the size of the matrix (M>N or M<N) and the difficulties associated with dealing with missing data.

Calculating EOFs from a data matrix

Consider a data set described by a data matrix $D$:

$$
D = \begin{pmatrix}
    d_{11} & \cdots & d_{1n} \\
    \vdots & \ddots & \vdots \\
    d_{m1} & \cdots & d_{mn}
\end{pmatrix}
$$

Each row represents the data at all times at position $x_j$ and each column represents the data at all positions at time $t_i$.

Observations of more than one variable at the same locations and time can augment the data matrix with additional rows, e.g. joint observations of temperature and salinity.

$$
d_{ij} = \begin{pmatrix}
    T_y \\
    S_y \\
    O_y
\end{pmatrix}
$$
The EOF method seeks to separate these data into the product of a set of time series and spatial patterns:

\[ d(x_i, t_j) = \sum_{k=1}^{M} \phi_k(x_i) a_k(t_j) \]

\( a_k(t) \) is the amplitude time series of the \( k^{th} \) mode (\( N \) elements)

\( \phi_k(x) \) is the spatial pattern (also called the “loading”) of the \( k^{th} \) mode (\( M \) elements)

If we make a matrix \( A \) with rows being the time series \( a_k(t) \)

and another matrix \( E \) with columns being the spatial patterns \( \phi_k(x) \)

then the summation over the modes above of the product of time series and spatial patterns is simply the matrix product:

\[ D = EA \]

**The EOF modes**

Fundamental to the EOF method is that the modes are *orthogonal*.

\[ \sum_{k=1}^{M} \phi_i(x_k) \phi_j(x_k) = \delta_{ij} \]

or

\[ \phi_i^T \phi_j = \delta_{ij} \]

There are an infinite set of possible choices for the \( \phi_i \) that could comprise an orthonormal spanning set of basis functions to describe the data matrix \( D \).

EOFs are determined uniquely among the possible choices by the constraint that the time amplitudes \( a_k(t) \) are uncorrelated over the sample data.

\[ \overline{a_i(t)a_j(t)} = \lambda_i \delta_{ij} \]

where the overbar denotes a time average.

\[ \overline{a_i^2(t)} = \frac{1}{N} \sum_{k=1}^{N} a_i^2(t_k) = \lambda_i \]

so \( \lambda_i \) is the variance of time series \( i \).

More compactly:

\[ a_i^T a_i = N \lambda_i \]
The important thing to note is that $\lambda_i$ is proportional to the variance of mode $i$. So we see that EOFs have something to do with the variance of the data matrix.

For convenience let’s just roll the factor of $N$ into $\lambda_i$, and notice that the matrix $A$ constructed from the time series will have the property

$$AA^T = \lambda I = L$$

where $L$ is a matrix that looks a lot like the identity matrix except with the values $\lambda_i$ down the diagonal.

Consider the data covariance matrix constructed from the covariance of each time series with every other. Each element would be:

$$C_{i,j} = \frac{1}{N} \sum_{k=1}^{N} d(x_i, t_k) d(x_j, t_k)$$

This is simply:

$$C = DD^T$$

except for the factor of $N$.

It follows that

$$DD^T = EA(EA)^T = EAA^TE^T = E(\lambda I)E^T = ELE^T = LEE^T$$

Right multiply this by $E$

$$DD^TE = LEE^TE = LE$$

or

$$CE = LE$$

where $C$ is the data covariance matrix.

If we consider what this equation states for just one column of $E$:

$$C\phi_i = \lambda_i \bar{\phi}_i$$

we recognize that $\lambda_i$ and the columns of $E$ are eigenvalues and eigenvectors of the covariance matrix.
The set of eigenvectors and associated eigenvalues represent a coordinate system in which the transformed covariance is diagonal, with the variability in these directions being uncorrelated with each other.

We can get the amplitude time series corresponding to a particular eigenvector by projecting the data on to that eigenvector via the vector inner product (i.e. the component of \( \mathbf{D} \) in the direction of \( \hat{\phi}_m \))

\[
a_p(t_j) = \phi_m^T \mathbf{D}
\]

because this is

\[
\sum_{k=1}^{M} \phi_p^T(x_k) d(x_k, t_j)
\]

This is easily done all at once as

\[
\mathbf{A} = \mathbf{E}^T \mathbf{D}
\]

because it follows that

\[
\mathbf{E} \mathbf{A} = \mathbf{E} \mathbf{E}^T \mathbf{D} = \mathbf{E} \mathbf{E}^T \mathbf{D} = \mathbf{D}
\]

which is our original definition of the factorization of the data matrix.

So the eigenfunctions of the covariance matrix give us an expansion that completely describe the original data.

Seeking the eigenvectors as the set of basis functions describing spatial modes was arbitrary. I could have sought an expansion in terms of the vectors \( a^T \).

Not surprisingly, \( \phi_i, \lambda_i, a_i^T \) are interrelated. \( \phi_i \) and \( a_i^T \) are solutions to different eigenproblems that have the same eigenvalues.

Consider a matrix \( \mathbf{B} \) constructed from the data matrix thus:

\[
\mathbf{B} = \begin{pmatrix}
\mathbf{0}_{N \times N} & \mathbf{D}^T \\
\mathbf{D}_{M \times N} & \mathbf{0}_{M \times M}
\end{pmatrix}
\]

with eigenvectors \( \mathbf{q} \) which satisfy

\[
\mathbf{B} \mathbf{q} = \gamma \mathbf{q}
\]

Consider just one of the eigenvectors:
This is two separate problems:

\[
\begin{pmatrix}
q_{i,1} \\
\vdots \\
q_{i,N}
\end{pmatrix} = \gamma_i
\]

and

\[
\begin{pmatrix}
q_{N+1,i} \\
\vdots \\
q_{N+M,i}
\end{pmatrix} = \gamma_i
\]

which are

\[Dv_i = \gamma_i u_i\]

\[D^Tu_i = \gamma_i v_i\]

Multiply the respective equations by \(D\) and its transpose:

\[D^TDv_i = \gamma_i D^Tu_i = \gamma_i^2 v_i\]

\[DD^Tu_i = \gamma_i Dv_i = \gamma_i^2 u_i\]

So for two different matrices we have different eigenvector solutions, but the same eigenvalues.

Recognize the similarity between:

\[DD^Tu_i = \gamma_i^2 u_i\]

and

\[DD^T\phi_i = \lambda_i \phi_i\]
The vectors $u$ are analogous to the EOFs of the covariance matrix, and the vectors $v$ are analogous to the time series vectors $a$.

If one of $M$ or $N$ is much smaller than the other, we need only solve the smaller eigenvalue problem.

It can be shown that $u$ and $v$ are the singular vectors, and $\gamma_i$ are the singular values, of the data matrix.

Assembling the vectors into matrices we have

$$D = USV^T$$

The matrix $S$ has the singular values down the diagonal and is padded out with zeros to make it $M \times N$.

For example:

$$S = \begin{pmatrix}
\gamma_1 & 0 & 0 & 0 \\
0 & \gamma_2 & 0 & 0 \\
0 & 0 & \gamma_3 & 0 \\
\end{pmatrix}$$

The columns of $U$ and $V$ are orthogonal vectors because they are the eigenvectors of real symmetric matrices, so $U$ and $V$ are orthonormal matrices:

$$U^TU = UU^T = I$$
$$V^TV = VV^T = I$$

We can check the factorization of the data matrix:

$$DV = US$$

because we can switch the order of multiplication for the case of a diagonal matrix.

$$U^TDV = U^TUS$$
$$U^TDV = S$$

so we can say that $U$ and $V$ diagonalize $D$.

Also,
\[
\mathbf{DVV}^T = \mathbf{USV}^T \\
\mathbf{D} = \mathbf{USV}^T
\]

which is the definition of the Singular Value Decomposition.

If we order the eigenvalues such that

\[
\gamma_1 > \gamma_2 > \gamma_3 \ldots
\]

then the corresponding eigenvectors define directions in the initial coordinate space along which the maximum possible variance can be explained.

The total variance of the data is

\[
\frac{1}{M} \sum_{m=1}^{M} \left\{ \frac{1}{N} \sum_{j=1}^{N} d(x_m, t_j)^2 \right\}
\]

which it can be shown is proportional to the sum of the eigenvalues of the covariance matrix:

\[
\sum_{i=1}^{M} \lambda_i
\]

So how does the SVD factorization relate to the original eigenvalue problem:

\[
\mathbf{DD}^T = (\mathbf{USV}^T)(\mathbf{USV}^T)^T \\
\mathbf{C} = \mathbf{USV}^T \mathbf{VS}^T \mathbf{U}^T \\
= \mathbf{US}^2 \mathbf{U}^T \\
= \mathbf{US}^2 \mathbf{U}^T
\]

from which it follows that

\[
\mathbf{CU} = \mathbf{US}^2 = \mathbf{S}^2 \mathbf{U}
\]

which we can compare to

\[
\mathbf{CE} = \mathbf{LE}
\]

and thus the left singular vectors are the EOFs of the covariance matrix we considered previously, and the eigenvalues \( \lambda_i = s_i^2 \).

The amplitude time series are
\[ A = E^T D = U^T (USV^T) \]
\[ = SV^T \]

An important property of EOFs is that for any number of modes, \( K \), there is no other expansion of the data for fewer than \( K \) modes that captures more of the variance of the original data set. The EOF expansion for \( K \) modes is thus effectively a compression of the data.

The compressed data set is

\[ \hat{d}(x_i, t) = \sum_{k=1}^{K} \phi_k(x_i) a_k(t) \]

In practice, relatively few modes \((K \ll M)\) typically describe the majority of the data variance.

Higher modes with small variance generally have noisy amplitude (or loading) patterns that are dominated by short wavelengths.

EOFs can therefore be used as a filter to eliminate unwanted high frequencies of data variability that have no coherence across the spatial domain.

If the EOF spatial modes have gaps, for example because of a defective instrument in a network, then if the EOF loadings are filled by interpolation a reconstruction of the full data set can be achieved by matrix multiplication that should not introduce high frequency noise that might arise from more simple minded gap filling interpolations.

We can explore the SVD of a data matrix by example, by considering sea surface temperature variability in the equatorial Pacific Ocean.

See files in:

[http://marine.rutgers.edu/dmcs/ms615/jw_matlab/eof](http://marine.rutgers.edu/dmcs/ms615/jw_matlab/eof)