Weds
3-5-03

Properties of Eigenvectors

$Ax = \lambda x$

$(A - \lambda I)x = 0$

- The eigenvectors of a matrix form a linearly independent set
- If A $(M \times M)$ has full rank, the eigenvectors of A form a basis of R^M
  - If A is singular, with rank $k < M$, then the first k eigenvectors span the subspace of A & the last $k+1 : M$ eigenvectors span the nullspace of A.
    - If A is real & symmetric, the eigenvectors are not only linearly independent but they are orthogonal

Proof of the last statement: Let C be a real, symmetric matrix. Consider the eigenvectors e_k & e_j. Then:

$$Ce_j = \lambda_j e_j$$
$$Ce_k = \lambda_k e_k$$

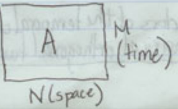
Transpose the top equation & multiply by e_k. Multiply the bottom equation by e_j^T. Recall: $(AB)^T = B^T A^T$

$$C = C^T \quad e_j^T C^T e_k = \lambda_j e_j^T e_k$$
$$-\left(e_j^T C e_k = \lambda_k e_j^T e_k\right)$$

$(\lambda_j - \lambda_k) e_j^T e_k = 0$ For distinct eigenvalues, $e_j^T e_k = 0$
$\Rightarrow$ the eigenvectors are orthogonal

Introduction to EOFs

Basic Premise: Say we have a data matrix with M observations of length N state vectors



We want to find a single $1 \times N$ state vector (e.g. spatial pattern) that explains the largest fraction of the total variance in the data matrix

Let e_1 be the vector that has the highest possible resemblance to the ensemble of state vectors in the data.

- The resemblance of e_1 to the state of the system is given by the mean projection of e_1 onto the ensemble of state vectors
- The projection is squared for the same reason that we squared the error in our estimate of the "goodness of fit" in linear regression.
- The vector e_1 is normalized to unit length $\|e_1\| = 1$ so that only the direction of e_1 impacts the projection
- Hence, we want to maximize the following:

$$\frac{1}{N}\left[{}_{1}\overset{N}{\boxed{e_1^T}} \; {}_{N}\overset{M}{\boxed{A^T}} \; {}_{M}\overset{N}{\boxed{A}} \; {}_{N}\boxed{e_1} \right] = \lambda$$

$$\boxed{\lambda = \frac{1}{N} e_1^T A^T A e_1}$$

<u>Note</u>: $C = \frac{1}{N} A^T A$

covariance matrix

$(e_1 e_1^T = 1$ by construction$)$

$\lambda = e_1^T C e_1 \xrightarrow[\text{as}]{\text{rewrite}} \boxed{C e_1 = \lambda e_1}$

This tells us that e_1 <u>must</u> be an eigenvector of C w/ corresponding eigenvalue (i.e. amplitude) λ.

Hence, we find e_1 by eigenanalyzing C & choosing the eigenvector that corresponds to the biggest eigen values

- The eigenvectors of the temporal covariance (dispersion) matrix are called the <u>empirical orthogonal functions</u> of A.

e.g. 1000 (space)

12 $\boxed{A}$ — A^TA yields a 1000X1000 matrix; AA^T yields a 12X12 matrix

The eigenvectors of AA^T are PCs of A. The EOFs are found by regressing the data A onto the PC.

- The EOFs & PCs are interchangeable with respect to matrix multiplication

<u>Use of Singular Value Decomposition (SVD) to calculate EOFs</u>

- Any real rectangular matrix can be represented as the product of 3 special matrices:

$$\boxed{A = U \Sigma V^T}$$

A is MXN
U is MXM
Σ is MXN
V is NXN

- The decomposition of A into U, Σ, & V satisfies the following equations:

$AA^T u_i = \sigma_i u_i$
$A^TA\, v_i = \sigma_i v_i$

*The columns in U & V are the eigenvectors of AA^T & A^TA, respectively. This means they correspond to the PCs/EOFs of A.

*The columns in U & V are orthogonal (if they are orthonormal then the amplitude of each EOF/PC pair is contained in Σ)

*The columns of U span M

*The columns of V span N

The matrix Σ can be viewed as comprising two matrices, depending on the relative sizes of M & N.

- If M>N, $\Sigma = \begin{bmatrix} D \\ 0 \end{bmatrix}$ where D is an NXN diagonal matrix & O is a (M-N)XN matrix of zeros.

rank of A

• The first q elements along the diagonal of D are the "singular values" the square root of the eigenvalues shared by A^TA & AA^T.

• The $q+1 \mapsto N$ elements of D are zeros & correspond to the nullspa
of A.

• In the case where M=N & A is full rank:

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_N \end{bmatrix}$$

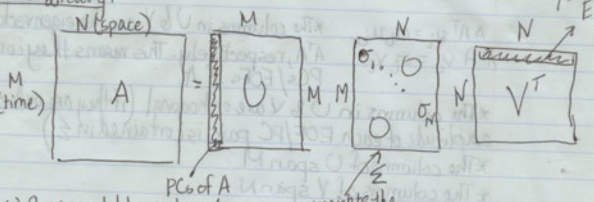
The largest possible rank of A is always $q \leq \min(M,N)$. Hence all colu in U & V where $i > q$ must lie in the nullspace of A.

<u>Example:</u> M>N & A is full rank. U spans R^M & V spans R^N.

• The N columns of V span the rowspace of A.

• The 1:N columns of U span the columnspace of A. The N+1:M columns in U lie in the nullspace of A.

<u>Schematically:</u>



weights the projection of each EOF

*U & Σ could be reduced to MXN & NXN w/ <u>no</u> impact on the results

• The pattern e_1 that explains the largest fraction of variance in the data matrix A is the eigenvector corresponding to the largest eigenvalue of C.

• The 2nd EOF is the eigenvector of C that corresponds to the second largest eigenvalue

<u>Consider the definition of eigenvalues/eigenvectors</u>

$Ax = \lambda x \Rightarrow CE = EL$

<u>where:</u> C is symmetric & real
E is the matrix of all eigenvectors e_i
L is a diagonal matrix with the eigenvalues (λ_i) along its diagonal

<u>In the case where C is the covariance matrix:</u>

Eigenanalysis transforms the dispersion matrix into a "coordinate space" where the "new" dispersion matrix is diagonal; i.e. the dispersion in the date is reordered along the diagonal.

• The coordinates in the transformed space are now orthogonal. Hence, eigenanalysis "diagonalizes" the dispersion matrix

• In the new coordinate space, <u>all</u> of the variance is found along the diagonal of the dispersion matrix. The fraction of the variance explained by eigenvector e_j is simply $\frac{\lambda_j}{\Sigma\lambda}$

Fri. 3-7-03 <u>Visualizing Coordinate Transformation via EOF Analysis</u>

Matrix A (100x2)

• Correlation between A(:,1) & A(:,2) is r = 0.7622. The first EOF of A is $[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$.

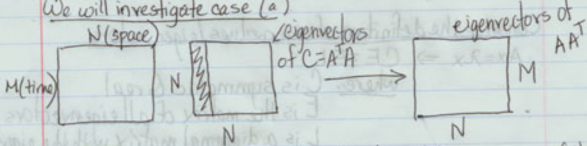
↳ fraction of variance explained is 88%.

EOF2 ⇒ $[-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}]$

Time series of the EOFs (called principal components)

- Found by either:
  - a) projecting the data onto the EOFs
  - b) computing the eigenvectors of the spatial dispersion matrix

We will investigate case (a)



- The projection yields a coefficient matrix of the expansion of the data in terms of the eigenvectors
  - The first PC is found by projecting the data onto EOF1, etc.

A quick summary of how to find EOFs using eigenanalysis of the temporal dispersion (i.e. covariance) matrix

Given: Data matrix A $(M \times N)$. Assume the time mean is zero at each grid point

- Calculate the dispersion relation $\frac{1}{N} A^T A$.
- Diagonalize (i.e. eigenanalyze) the dispersion matrix
- The eigenvector corresponding to the largest eigenvalue is EOF1
  - The eigenvector corresponding to the 2^{nd} largest eigenvalue is EOF2
- Find the PCs by projecting the data onto the EOFs.
- The fraction of variance explained by the n^{th} EOF/PC pair is $\lambda_n / \Sigma\lambda$

- No other linear combinations of k predictors can explain a larger fraction of variance than the first k PCs (EOFs).

- When the sample & structure dimensions are very different, what is the most efficient way to calculate the EOFs?

Mon.
3-17-03

EOF "Modes"

$A = U\Sigma V^T$ can be rewritten as: $A = \sum_{i=1}^{N} \sigma_i u_i v_i^T$ where each i can be thought of as an EOF "mode". Schematically:

$$A = \underbrace{\sigma_1 \underbrace{\begin{bmatrix} \vdots \\ u_1 \\ \vdots \end{bmatrix}}_{PC1} \underbrace{\begin{bmatrix} \cdots v_1^T \cdots \end{bmatrix}}_{EOF1\,(N\times1)}}_{\text{mode 1}} + \sigma_2 \underbrace{\begin{bmatrix} \vdots \\ u_2 \\ \vdots \end{bmatrix}}_{PC2} \underbrace{\begin{bmatrix} \cdots v_2^T \cdots \end{bmatrix}}_{EOF2} + \ldots$$

Equivalence of SVD to Diagonalication

$A = U\Sigma V^T$ Square both sides:

$A^TA = V\underbrace{\Sigma^T \underbrace{U^TU}_{\equiv I} \Sigma}_{\equiv \Sigma^2} V^T$ U, V are orthonormal

$\boxed{CV = V\Sigma^2}$ Eigenvalue Problem

Mon.
3-24-03

Summary of Finding EOFs via SVD

1) Decompose matrix A such that $A = U\Sigma V^T$. This is done using eigenanalysis, but most software packages offered canned SVD routines

2) For $A_{M\times N}$, if M is the sample space (e.g. time), then the EOFs (spatial patterns) to the columns in V (the eigenvectors of A^TA). The PCs are the columns in U (eigenvectors of AA^T).

3) The fraction of variance explained by the n^{th} EOF/PC pair is: $\dfrac{\sigma_n^2}{\Sigma\sigma^2}$

Presentation of EOFs/PCs

• Whether the EOFs or PCs or both are shown depends on which domain yields interesting structures. In general, we are interested in patterns in the spatial domain.

• Problem: The EOFs are generally dimensionless. It is useful to display the EOFs in the physical units of the data being analyzed. The standard way to do this is to regress the data onto the standardized values of the PC.

• Nice to plot variance (standard deviation) of the fields along with the EOF → see where the pattern matches the data.

In the case of eigenanalysis of the covariance matrix

1) For the dispersion matrix $A^T A$, where $A_{M \times N}$, C is $N \times N$.
time ↗ ↖ space

• project the data onto e_1 to get the first PC.

$$A_{M\times N} \cdot e_{1(N\times 1)} \Rightarrow \underbrace{z_{1(M\times 1)}}_{PC}$$

2) Standardize z_1 $\quad \hat{z}_1 = \frac{z_1 - \bar{z}_1}{\sqrt{z_1'^2}}$

3) Regress the data onto $\hat{z}$.

$$\hat{z}_1^T \cdot A_{M\times N} \Rightarrow \hat{e}_{1(1\times N)}$$ ← divide by M to get units of regression

In the case of SVD

• By construction, the PCs & EOFs are orthonormal. If $A_{M\times N}$ (M is time):

$$A = U \Sigma V^T$$
$$U^T A = U^T U \Sigma V^T$$
$$\boxed{U^T A = \Sigma V^T}$$ → multiplying the orthonormal EOFs by the corresponding singular values is equivalent to projecting the data onto the orthonormal PCs.

- The orthonormal EOFs in V (or U) multiplied by Σ have physical units, BUT do not have the same amplitude as the data regressed onto the standardized PCs. This is because the orthonormal PCs are not necessarily standardized.
  - The singular values are really only useful for assessing % of variance explained.

EOFs based on the correlation matrix

- Sometimes it is desirable to remove the amplitude from the data before computing the EOFs.

1) If the state vector is a combination of parameters with different units.
2) If the variance of the state vectors varies greatly from space to space.

In general, when using EOF analysis on a geophysical field, retain the amplitude of the data.

a) EOFs of the unnormalized data generally explain more variance.
b) The centers of unnormalized EOFs are shifted towards regions of high variance. Since gradients in variance are inextricably linked to the spatial structure of the field, it seems unphysical to eliminate them.

Weds. 3-26-03

How many EOFs should be retained?

White noise → no memory; equal power @ all periods

- The EOFs of white noise are degenerate (the eigenvalues are not distinct). The off-diagonal elements of the dispersion matrix are zero. The diagonal elements are equal but for sampling fluctuations. The eigenvectors of the identity matrix $I_{N \times N}$ has N orthogonal eigenvectors but only one eigenvalue: 1.

For red noise: The covariances between nearby elements in the dispersion matrix fall off exponentially.

• EOFs of red noise are generally "pleasing" to look at (they reflect sines & cosines) but are not necessarily physical. The rate at which eigenvalues drop off depends on the redness of the data.

· The orthogonality constraint on EOFs imposes a strong limitation on the shape of the EOFs. Successive EOFs must be orthogonal to higher EOFs.

· The statistical significance (robustness) of EOFs can be assessed using the t-statistic based on sample size & the number of variables in the state space (the # of eigenvalues).

· North et al. (1982) show that the significance of EOFs is a function of separation between eigenvalues:

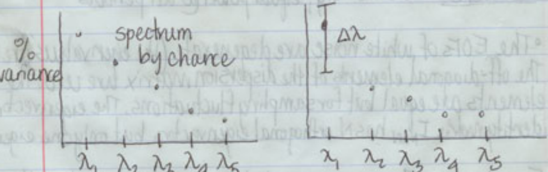
sampling error

$$\boxed{\Delta\lambda = \lambda\sqrt{\frac{2}{N'}}}$$

number of degrees of freedom

The λ falls between $\lambda+\Delta\lambda$ & $\lambda-\Delta\lambda$

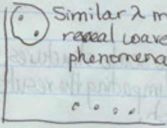
· If the confidence ranges for the two EOFs overlap, then they are not considered "distinct", at least as far as North et al. criterion is concerned



Comments on significance

- In general, N is hard to test. To test the significance, subdivide the data, e.g. w/ Monte Carlo tests.
  - Plot spatial correlations w/ each EOF (EOF1, EOF2, etc) & see how much they vary.

Special Case



- Are the results explainable in terms of theory? Are the results robust to change in size of domain, independent data, etc.

Preparing The Data

- If the time mean in the data is zero, the PCs will have zero mean.
- If you remove the spatial mean in the data, the EOFs will have zero mean.
- Common to remove the time mean from the data s.t. A^TA corresponds to the covariance matrix.
  - If you don't remove the time mean, the axis of the PCs will pass through the origin, not the centroid of the data.
- If you are interested in anomalies, you would remove the seasonal/diurnal cycle.
- It is less common to remove the spatial mean → want to allow the EOFs to have the same sign everywhere

Two exceptions

1) You are interested in gradients (eg. ∇T)
2) You are interested in the eddy component of the data (remove the zonal mean)

Weighting the data

- If the data are gridded, need to weight according to the grid size. For latitude/longitude grids, weight by $\sqrt{\cos\varphi}$. ($\sqrt{\cos\varphi}$ since we are eigenanalyzing A^TA).

- Note that the results yield EOFs with the weighting built in. Show the regression maps based on the PCs.

Condensing the data

- Since EOFs typically reflect large scale structures, you can almost always condense the data w/o impacting the results
  - Interpolate to a broader mesh.
  - Use Fourier Analysis.

Missing Data

- Most programs will not eigenanalyze or perform SVD w/ missing values.
  - ~~Possible solutions~~
    - Interpolate raw data field.
    - Estimate the covariances w/ available data so that A^TA has no missing values.

Mon
3-31-03

Rotated EOF Analysis (Richman 1986)

Basic idea: The orthogonality constraint of EOFs often yield structures that are artificially wave-like.

• EOFs tend to have the largest possible spatial scale since they seek to maximize the variance explained by one spatial pattern.

• EOF patterns can be simplified by rotating the first N EOFs (Rotation ⇒ linear combinations)

• The criteria that have been advocated for obtaining the optimal rotations can be separated into two types: orthogonal & oblique rotations.

orthogonal: preserve orthogonality of the PCs — most commonly seen in the literature.

oblique: do not preserve the orthogonality of the PCs

*In both cases, the orthogonality of the EOFs is relaxed.

General Mathematics

• REOF analysis is the coordinate transformation from A (eofs) to B (REOFs) by means of an invertible matrix $R_{N\times N}$ (where N is the number of EOFs being rotated & hence the number of REOFs).

$$B = RE$$

EOFs lie in the rowspace

• For each rotated vector b_i: $b_i = \sum_{j=1}^{N} r_{ij} e_j$

e.g. $b_1 = r_{1,1}e_1 + r_{1,2}e_2 + r_{1,3}e_3 + \dots + r_{1,N}e_N$

$b_2 = r_{2,1}e_1 + r_{2,2}e_2 + r_{2,3}e_3 + \dots + r_{2,N}e_N$

• The matrix R is chosen s.t. a particular function S is maximized. Generally the idea is to seek the simplest possible structures in B.

This means we must:

1) have a reason to expect simple structures
2) define simplicitly mathematically

- Simplicity is generally defined as when most elements in B are of order 1. (absolute value) or 0, but not in between.

- The varimax criterion is the most commonly used criterion for orthogonal rotation. In this case, simplicity is defined as the variance of the squared loadings.
  - Hence, for REOF b_i, we want to maximize the function:

$$S_p^2 = \frac{1}{M}\sum_{j=1}^{M}\left[b_{ij}^2 - \bar{b_i^2}\right]^2$$

Maximize the amplitudes

where M is the number of elements in REOF b_i (i.e. the number of gridpoints in the data) &:

$$\bar{b_i^2} = \frac{1}{M}\sum_{j=1}^{M} b_{ij}^2$$

- When s_p^2 is maximized, the loadings in each b tend to be 0 or 1.
  1) The simplification ignores any weighting to the input EOFs. The input EOFs are often weighted by the square roots of their respective eigenvalues. Leading EOFs contribute more to the varimax criterion than lower EOFs.
  2) REOFs generally yield regioned patterns that resemble 1 point correlation maps.
  3) If the number of EOFs = number of grid points, then the REOFs degenerate into localized "bulls-eyes" scattered around the domain.

· If N is very small, the results are very sensitive to the choice of N. N~10 is a good number.

4) The uniqueness of REOFs does not derive from the separation of variance explained but from the role in explaining local variance.

· REOFs may be useful when $\lambda_1 \sim \lambda_2 \Rightarrow$ the pattern is some linear combination of EOF1 & EOF2.

Summary

· The jury is still out on REOFs

1) The choice of N is arbitrary.
2) Small changes in R can yield very different results.
3) The results are sensitive to the weighting applied to the input EOFs.
4) REOFs add another level of complexity to the analysis. They are more difficult to reproduce.

Weds. 4-2-03

Methods for Finding Coupled Patterns

· MCA → Maximum Covariance Analysis (SVD analysis)

more preferable name...

· References Bretherton et. al. 1992 & Wallace et al. 1992

· Extended EOF Analysis

· Canonical Correlation Analysis (CCA)

1) MCA

· Recall the use of the matrix operation SVD

a) SVD of the data matrix $A_{M\times N}$ (M is time; N is space)

$A = U \Sigma V^T$

– the columns of U & V are the orthonormal eigenvectors of AA^T & A^TA, respectfully.

- the diagonal elements of Σ are the singular values (the square roots of the eigenvalues)

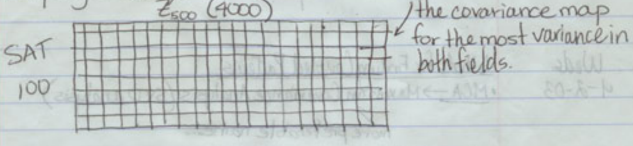
b) SVD of the dispersion matrix $C_{N\times N} = A^T A$

$A = U \Sigma V^T$

$C = V \Sigma V^T$ (all $N\times N$)

- The columns of V are the orthonormal eigenvectors of C (you might expect the columns in V to be the orthonormal eigenvectors of $C^T C$. They are. The eigenvectors of C & $C^T C$ are the same).
- In the case of SVD($A^T A$), the eigenvectors are the same as in SVD(A). The Σ in SVD($A^T A$) correspond to the eigenvalues of $A^T A$ & $A A^T$.

- In MCA, you apply SVD to the dispersion matrix between two different data sets (data sets with different state vectors but a common sampling domain).



- The results give you patterns in one data set whose time series explain the largest fraction of variance in the other data set & vice versa.
- MCA can be used to isolate coupling between two physical fields.

<u>The Mathematics of MCA</u>

- Suppose we have two data sets: The "left" field $X_{M\times N}$ & the "right" field $Y_{M\times L}$. M is the shared sampling dimension & N & L are the structure dimensions.

First compute the NXL dispersion matrix.

$$C_{XY} = \frac{1}{M} X^T Y$$

Second: Apply SVD to C_{XY}

$C_{XY} = U \Sigma V^T$ → The columns in $U_{N \times N}$ (referred to as singular vectors) span the column space of C_{XY}. They correspond to structures in X

→ The columns in $V_{L \times L}$ correspond to structures in Y.

Third: The time series are found by projecting the original data sets onto the respective "singular vectors" in U & V.

- X is projected onto U.
- Y is projected onto V.

- The expansion coefficient time series are not mutually orthogonal. The diagonal elements in Σ have units of covariance.

Fourth: Display your results. There are two types of maps are commonly used.

Homogeneous Regression Maps: Regress (or correlate) the input data in the left field (X) onto the expansion coefficient time series for the same field (X).

→ These maps have high amplitude.

Heterogeneous Regression Maps: Regress the input data in the left field (X) onto the expansion coefficient time series of the right field (Y) & vice versa.

- These maps are a more stringent test of coupling.

Significance $\Rightarrow$ Does one mode "stand out" from the others?

- Fraction of the squared covariance explained.
- The squared covariance for one mode divided by the sum of the squared covariances.

$$\Rightarrow \frac{\sigma_i^2}{\Sigma \sigma_i^2}$$, where σ is the diagonal element of Σ.

$\Rightarrow$ How well coupled are the data in general?

$\rightarrow$ Root mean squared covariance is the sum of the diagonal elements in Σ.

- By itself, this number is not very informative. Hence, it is useful to normalize the number by the sum of the product of the variance of the left & right fields.

$$RMSC = \frac{\sum_{i=1}^{N}\sum_{j=1}^{L}\left(\overline{x_i y_j}\right)^2}{\sum_{i=1}^{N}\overline{x_i^2}\sum_{j=1}^{L}\overline{y_j^2}}.$$

Fri. 4-4-03

Extended EOF Analysis

- Suppose we have two data sets: X_{MxN} & Y_{MxL}
  - EEOF analysis is simply EOF analysis of the combined dataset $A_{Mx(N+L)} = [X\ Y]$
- Easiest to visualize the relationship between EEOF analysis & MCA via the covariance matrix.
  - The covariance matrix for $A_{Mx(N+L)} = [X\ Y]$ is

	N	L	
N	X^TX (the covariance matrix used for EOF)	X^TY (the covariance matrix for MCA)	N
L	Y^TX	Y^TY	L
	N	L	

$C = A^TA$

Hence, EEOF is analogous to MCA, but it also contains information about the structures that dominate independent variability in X & Y.

Canonical Correlation Analysis

Basic Idea: Perform EOF & MCA in sequence. The MCA is performed on a truncated set of EOFs which explain a large fraction of the variance in the original data set.

- It is argued that CCA is less likely to yield patterns that are correlated by chance; truncation reduces the "noise" in the original data.

Truncation of Data

→ Data sets X & Y are truncated to N EOFs each. The choice of N reflects a tradeoff between adding noise to the analysis and retaining variance explained.

↳ The PCs are standardized before computing the dispersion matrix.

→ MCA is performed on the resulting correlation matrix.

→ In practice, CCA yields very similar results to MCA, but takes more time.

→ CCA & MCA are subject to sampling variations.

→ CCA is used for statistical prediction

→ CCA, MCA, EEOF, & EOF are all excellent tools for exploring data.