3) The fraction of variance explained by the i^{th} EOF / PC pair is:

$$\frac{\sigma_i'^2}{\sum_{i=1}^{N} \sigma_i'^2} \quad \text{or} \quad \frac{\lambda_i'}{\sum_{i=1}^{N} \lambda_i'}$$

Presentation of EOFs

Whether the EOFs or PCs are shown depends on which domain yields interesting structure.

Problem with looking directly at EOFs is that they are normalized, so don't get a sense of physical units.

Common way to show the EOF spatial patterns is to regress the data (i.e A matrix) onto the standarized PC time series.

Steps

1) Get the PC { via SVD
   projecting A onto EOFs via diag. method

2) Standardize the PC so it has a variance = 1 (unit variance)

$$PC = \frac{PC}{\sqrt{\overline{PC'^2}}} \rightarrow \text{std. dev. of PC.}$$

3) Regress the data onto the $\hat{PC}$ (normalized) Yields the EOF scaled in actual units.

* Assumes a linear regression.

→ So don't show the "raw" EOF

In the case of SVD.

$$A = U \Sigma V^T$$

Multiply by U^T (i.e. project PCs)

$$U^T A = \underbrace{U^T U}_{\substack{\text{yields} \\ \text{identity matrix}}} \Sigma V^T$$

$$U^T A = \Sigma V^T$$

Multiplying the orthonormal EOFs by corresponding singular values equivalent to projecting data on the orthonormal PCs.

Caveats

→ "Orthonormal" is not the same as standardized

→ σ^2 (or λ), that is the eigenvalues, are useful for telling % variance explained, but <u>not</u> for scaling EOFs.

EOFs using the correlation matrix

Recall get the correlation matrix instead of the covariance matrix ($A^T A$) if the time series of the data are standardized before the analysis, typically assuming a normal distribution.

$$z = \frac{x - \mu}{\sigma}$$

μ = time avg. over M time intervals

σ = std. deviation over M time intervals.

Do this at each point.

Reasons to do:

1) if the state vector is comprised of different parameters
2) if the variance of the state vector varies greatly from one location to another.

=> Generally NOT the case when working with space - time data.

In general, suggest retaining amplitude in data because:

1) EOFs of unnormalized data explain more variance.
2) Centers of unnormalized EOFs are shifted from actual EOFs. (i.e. spatial pattern may change).

*The spatial patterns and gradients in variance matter, as they're linked to the spatial structure of the field.

Possible disadvantage: Wipe out ~~regions~~ regions of small variance (e.g. mid-latitudes vs. tropics)

How many EOFs should be retained?

"White noise" = time series with equal power at all periods (i.e. no dominant time frequency in the data) and no spatial structure.

In such a case, all the eigenvalues (λ) would be equal to one. Therefore no one pattern would be any more significant than another pattern. ~~[illegible]~~ Off diagonal ~~[illegible]~~ elements in the covariance matrix $= 0$. because no spatial structure.

The processes in the atmosphere, and most geophysical systems, are "red noise."

"Red noise": Because data are autocorrelated in time, the largest space scales and lowest frequencies tend to explain more of the variance than smaller scales and higher frequencies.

In the case of red noise:

→ Correlation between nearby elements in the covariance matrix will fall off exponentially

→ Eigenvalues will fall off exponentially (1st order linear autoregressive process).

Don't be fooled! EOF caveats:

→ EOFs of red noise are generally pleasing to look at and seem like "real" spatial structure (just reflect autocorrelation)

→ Orthogonality constraints impact EOF structure

→ Higher order EOFs generally reflect mathematical constraints, but are not necessarily physical!!

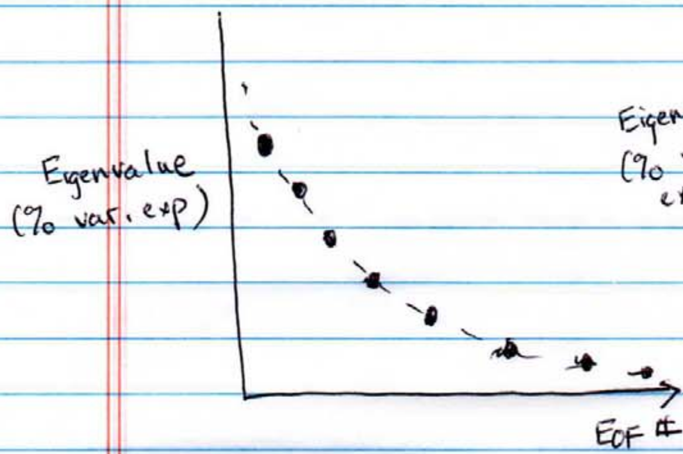
(e.g. 10th EOF doesn't mean anything!!)

North test for EOF significance

North et al. (1982) showed that the significance of EOFs is a function of the degree of separation between eigenvalues (λ)

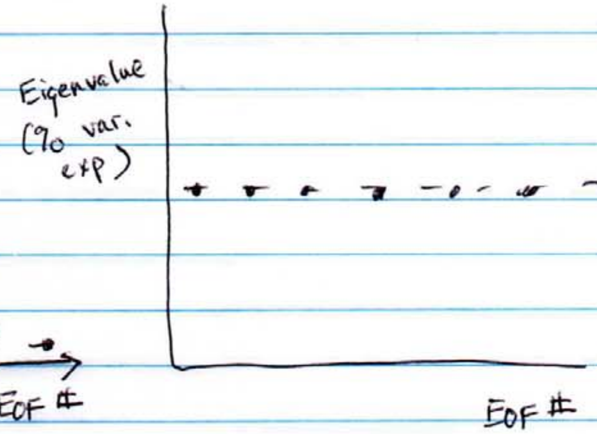
Basic idea: The more separated the EOFs, eigenvalues, more likely to be statistically significant.

Red noise



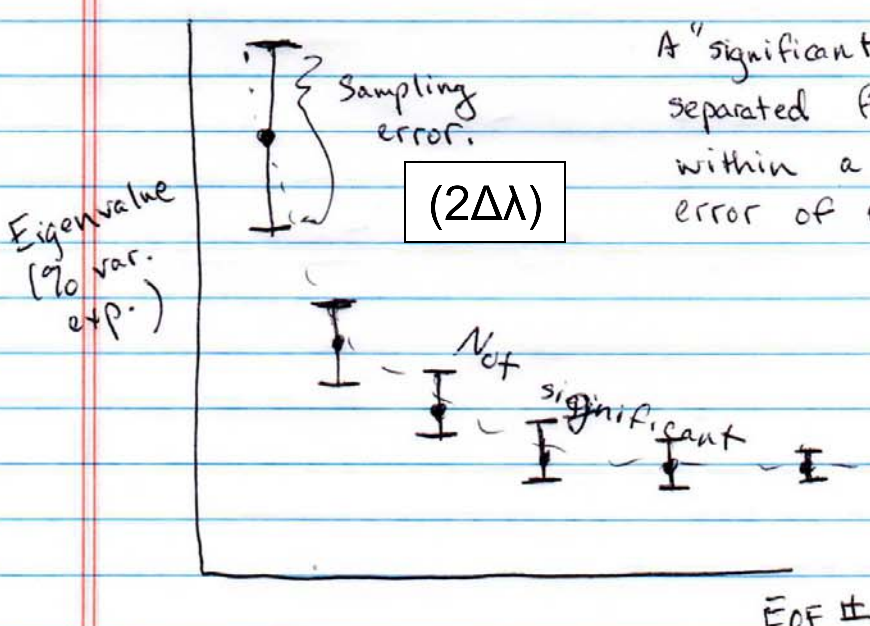
Eigenvalues decrease, but not well separated.

White noise



Eigenvalues are the same for all EOFs.

Significant EOFs – "Well separated"



A "significant" EOF is separated from the others within a sampling error of it's eigenvalue

Calculating the sampling error $(\Delta\lambda)$:

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

λ = eigenvalue

N^* = effective sample size.

→ Gives 95% confidence error in estimation of eigenvalues.

<u>A note on N^*</u>

N^* = Effective sample size.

- In practice, this is very hard to determine!
- As Wilk's notes on p. 492 "an appropriate modification to the effective sample size adjustment appears not to have been published"

Wilk's suggestion: Use lag-1 autocorrelation.

$$n' = n\left(\frac{(1-\rho_1^2)}{(1+\rho_1^2)}\right)^2$$

I would assume calculate this for each point in domain and then take the domain average, but Hartmann, Wilks, and von Storch are not clear about it!!

My suggestion:

Whenever doing EOFs, first look~~ing~~ at a spatial map of the lag-1 autocorrelation. If its large, increase the time interval by averaging the data until ρ_1 is sufficiently small to be negligible $\rightarrow$ then it's okay to assume data are roughly independent.

Homework example: It's probably okay to assume 500-mb heights in the NH are pretty much independent year to year, but not month by month or day by day.

For this course, we'll always assume independent data for EOF analysis, hence $N^* =$ sample size N.

Preparing data for EOF / PC Analysis

Time mean in data = 0 → PCs have zero mean
Spatial mean = 0 → EOFs have zero mean.

1) To get covariance matrix → remove time mean from data

For each point $x' = x - \mu$

2) To get correlation matrix → remove time mean from data and divide by the std. deviation

For each point $z = \frac{x - \mu}{\sigma}$

if don't remove the time mean, the axis of PCs will pass through the origin, not the centroid of the data.

If time means large, EOFs dominated by mean field.

Less common to remove the spatial mean
Examples of when to do:
→ Interested in steep gradients
→ Eddy variance

Data weighting

If data are gridded, need to weight according to the grid size.

For lat/lon grids:
Weight data by the the square root of the cosine of latitude (yields cov. matrix weighted by cosine of latitude)

Result: EOFs lie in "weighted" space.

Generally show raw data then regressed on corresponding PC time series.

Final notes on basic EOF/PC analysis

- EOFs by construction typically reflect large patterns.
- Most computer programs will NOT eigenanalyze or perform SVD with missing data.
- Covariance matrix is estimated using data at each point for each time. If missing data - need to omit that point.
- Remember that if you do SVD on covariance matrix, the singular values (σ) must be squared to get the eigenvalues.