goals of EOF analysis

1. economy of description → reduce data → ~~discard~~ reveal redundancies
2. orthogonalization → transform into a set of variables that are mutually uncorrelated, linear independent

"When a large number of measurements are available, it is natural to enquire wheater they could be replaced by a fewer number of the measurements or of their functions, without loss of much information, for convenience in the analysis and in the interpretation of data" Rao (1964)

2009

great Indian statistician

vector space A (measured variables) vector space B (statistical variables)

$$\begin{pmatrix} \text{currents @ } x_1 \\ \text{currents @ } x_2 \\ \text{currents @ } x_3 \\ \text{wind @ } x_4 \\ \text{wind @ } x_5 \\ \text{sealevel @ } x_6 \\ \text{sealevel @ } x_7 \\ \text{rainfall @ } x_8 \\ \vdots \end{pmatrix} = \begin{pmatrix} u(t,\vec{x}_1) \\ u(t,\vec{x}_2) \\ u(t,\vec{x}_3) \\ W(t,\vec{x}_4) \\ W(t,\vec{x}_5) \\ \eta(t,\vec{x}_6) \\ \eta(t,\vec{x}_7) \\ R(t,\vec{x}_8) \\ \vdots \end{pmatrix} = \begin{pmatrix} a_{11}(\vec{x}_1) \\ a_{12}(\vec{x}_2) \\ a_{13}(\vec{x}_3) \\ a_{14}(\vec{x}_4) \\ a_{15}(\vec{x}_5) \\ a_{16}(\vec{x}_6) \\ a_{17}(\vec{x}_7) \\ a_{18}(\vec{x}_8) \\ \vdots \end{pmatrix} \cdot Z_1(t) + \begin{pmatrix} a_{21}(x_1) \\ a_{22}(x_2) \\ a_{23}(x_3) \\ a_{24}(x_4) \\ a_{25}(x_5) \\ a_{26}(x_6) \\ a_{27}(x_7) \\ a_{28}(x_8) \\ \vdots \end{pmatrix} \cdot Z_2(t)$$

examples of NOAA altimeter data Pacific
Meenders + Chant (2000)

~~explain~~ "mode 1 explains 50% of the variance" "mode 2 explains 30% of variance"

need to interpret statistical modes physically ex.
→ not always easy

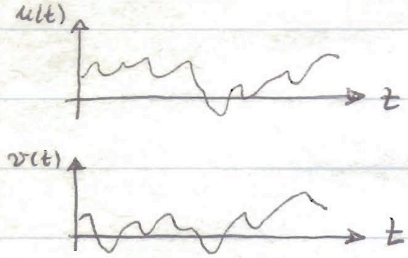
+ higher "modes"

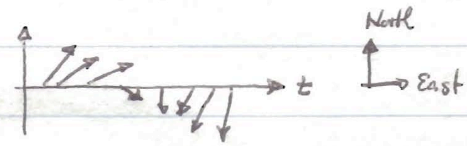
end #19

start class # 20

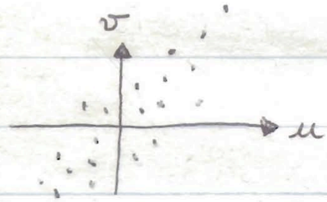
example: $M=2$ $\quad u_1(t) = u_1(t, x_1) \equiv u(t)$ east component

$u_2(t) = u_2(t, x_2) \equiv v(t)$ north component of velocity





time series of a vector (velocity) with $M=2$ components



Let's do an EOF of these and see what we will get

(a) compute the cross-covariance matrix

- remove mean $\quad u' = u - \langle \bar{u} \rangle \qquad \langle \bar{u} \rangle = \frac{1}{K} \sum_{\ell=1}^{N} u(t_k)$

$v' = v - \langle \bar{v} \rangle$

$$\langle \cdot \rangle = \frac{1}{K} \sum_{\ell=1}^{N} \cdot (t_\ell)$$

- $$\begin{pmatrix} \langle u'u' \rangle & \langle u'v' \rangle \\ \langle v'u' \rangle & \langle v'v' \rangle \end{pmatrix} = R_{ij} \quad \text{Reynold's stress tensor}$$

(b) find eigenvalues of the cross-covariance matrix (Reynold's stress tensor), i.e.

$$\begin{pmatrix} \langle u'u' \rangle & \langle u'v' \rangle \\ \langle v'u' \rangle & \langle v'v' \rangle \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \lambda \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

where λ is an eigenvalue and (e_1, e_2) is an eigenvector

$$\begin{pmatrix} \langle u'u' \rangle - \lambda & \langle u'v' \rangle \\ \langle v'u' \rangle & \langle v'v' \rangle - \lambda \end{pmatrix} \begin{pmatrix} e_1 \\ \bar{e}_2 \end{pmatrix} = 0$$

Cramer's rule states that solutions to linear equations are the ratio of 2 determinants, the denominator is the determinant of the matrix and the numerator is the matrix with the i-th column replaced by the vector on the right hand side, i.e.,

$$\begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 38 \\ 110 \end{pmatrix}$$

$$x_1 = \frac{\det\begin{pmatrix} 38 & 10 \\ 110 & 30 \end{pmatrix}}{\det\begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix}} = \frac{38 \cdot 30 - 10 \cdot 110}{4 \cdot 30 - 10 \cdot 10} = \frac{40}{20} = 2$$

$$x_2 = \frac{\det\begin{pmatrix} 4 & 38 \\ 10 & 110 \end{pmatrix}}{\det\begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix}} = \frac{4 \cdot 110 - 10 \cdot 38}{4 \cdot 30 - 10 \cdot 10} = \frac{60}{20} = 3$$

Apply this to our eigenvalue problem

$$e_1 = \frac{\det\begin{pmatrix} 0 & R_{12} \\ 0 & R_{22} \end{pmatrix}}{\det\begin{pmatrix} R_{11}-\lambda & R_{12} \\ R_{21} & R_{22}-\lambda \end{pmatrix}} = \frac{0}{(R_{11}-\lambda)(R_{22}-\lambda) - R_{12}R_{21}}$$

$$e_2 = \frac{\det\begin{pmatrix} R_{11} & 0 \\ R_{21} & 0 \end{pmatrix}}{\det\begin{pmatrix} R_{11}-\lambda & R_{12} \\ R_{21} & R_{22}-\lambda \end{pmatrix}} = \frac{0}{(R_{11}-\lambda)(R_{22}-\lambda) - R_{12}R_{21}}$$

For nontrivial, i.e. $(e_1, e_2) \neq 0$ solutions we need

$$(R_{11} - \lambda)(R_{22} - \lambda) - R_{12} R_{21} = 0$$

corrections Mar. 8, 2002

$$R_{11} R_{22} - \lambda R_{22} - \lambda R_{11} \overset{+}{\cancel{-}} \lambda^2 - R_{12} R_{21} = 0$$

$$\lambda^2 \overset{-}{\cancel{+}} (R_{11} + R_{22}) \lambda \overset{-}{\cancel{+}} (R_{12} R_{21} - R_{11} R_{22}) = 0$$

$$\lambda_{1/2} = \overset{+}{\cancel{-}} \frac{(R_{11} + R_{22})}{2} \pm \sqrt{\frac{(R_{11} + R_{22})^2}{4} \overset{+}{\cancel{-}} (R_{12} R_{21} - R_{11} R_{22})}$$

generally complex, however, it can be shown that for symmetric $(R_{12} = R_{21})$ and Hermitean matrices $(R_{12} = R_{21}^*)$ the eigenvalues are real.

Let's take $R_{ij} = \begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix}$ as before

then

$$\lambda_1 = 33.4$$

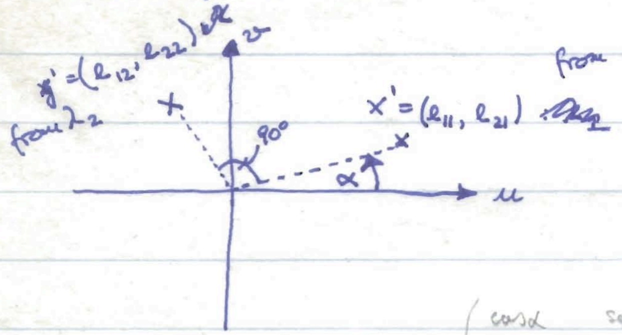
$$\lambda_2 = 0.6$$

note that $\lambda_1 + \lambda_2 = R_{11} + R_{22}$

In order to get the eigenvectors (e_1, e_2) we need to solve for each eigenvalue the linear equations

$$\begin{pmatrix} R_{12} - \lambda_m & R_{12} \\ R_{21} & R_{22} - \lambda_m \end{pmatrix} \begin{pmatrix} e_{1m} \\ e_{2m} \end{pmatrix} = 0 \qquad m = 1,2$$

What does this now mean, how to interprete the eigenvectors and eigenvalues?



that's a rotation α of the co-ordinate system.

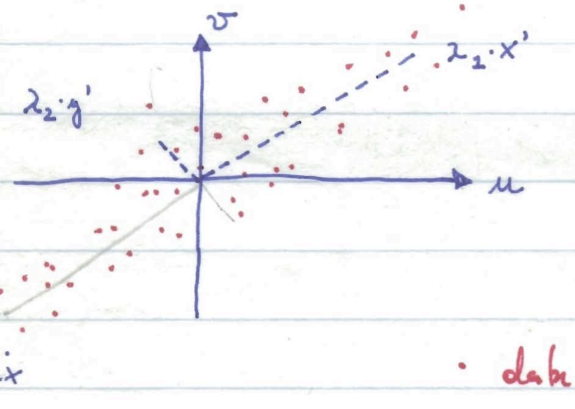
$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} A_1'(t) \\ A_2'(t) \end{pmatrix}$$

i.e., the eigenfunctions are $\{\sin\alpha, \cos\alpha\}$

where $A_1(t)$ and $A_2(t)$ are the "orthogonal" timeseries of velocities in the (x', y') frame

the eigenvectors define a "new" base, i.e., they rotate the co-ordinates from the (u, v) to the (x', y') system. For symmetric matrices the new base vectors are orthogonal, i.e., they are linear independent, i.e., they are mutually uncorrelated.

the eigenvalues represent the length, magnitude, relative importance of the respective "base" vectors which is closely related to the variance as $R_{11} + R_{22} = \lambda_1 + \lambda_2$



· data

that's an ellipse with a major and a minor axis $\lambda_1 x'$ and $\lambda_2 y'$, respectively

in the "new" system the covariance matrix is $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$

$\Rightarrow$ principal components of variability (u and v are correlated)
orthogonal base representing uncorrelated variability