13 Empirical Orthogonal Functions

13.0.0 Overview. In this chapter we present a multivariate analysis technique that is to derive the dominant patterns of variability from a statistical field (a random vector, usually indexed by location in space). *Principal Component Analysis*, or *Empirical Orthogonal Function* (EOF) Analysis as it is called in the Earth Sciences, was described by Pearson [309] in 1902 and by Hotelling [186] in 1935. EOF analysis was introduced into meteorology by Lorenz [259] in 1956.

Concepts in linear algebra that are needed to read this chapter (linear bases, matrix properties, eigenanalysis and singular value decomposition) are offered in Appendix B. Empirical Orthogonal Functions are formally defined in Section 13.1. Techniques for estimating EOFs, eigenvalues, and EOF coefficients are explained in Section 13.2. We discuss the quality of estimates in Section 13.3. Several EOF analyses of climate-related problems are given as examples in Section 13.4. *Rotated EOFs*[1] are dealt with in Section 13.5. Finally, a time series analysis technique called *Singular Systems Analysis*, which uses the same mathematics as EOF analysis, is introduced in Section 13.6.

An alternative introduction to EOFs is given by von Storch [387].

13.0.1 Introductory Example:[2] Daily Profile of Geopotential Height at Berlin. To motivate the concept of Empirical Orthogonal Functions we consider a time series of daily geopotential height profiles as obtained by radiosonde at Berlin (Germany) (Fraedrich and Dümmel [125]). A total of 1080 observations are available in each winter (NDJF) season: 120 days times 9 vertical levels between 950 hPa and 300 hPa. Thus, in a 20-year data set we have 21 600 observations at our disposal to describe the statistics of the geopotential height at Berlin in winter.

The mean state can be estimated by computing the mean value at each level. But how should we describe the variability? One way would be to compute the standard deviation at each level and to plot it in the vertical. However such a profile does not tell us how the variations are correlated in the vertical. For example, are we likely to observe a positive anomaly (i.e., a positive deviation from the mean profile) at 300 hPa and at 950 hPa at the same time?

EOF analysis is a technique that is used to identify patterns of simultaneous variation. To demonstrate the concept we let $\vec{\mathbf{x}}_t$ represent the $m = 9$ level geopotential height profile observed at time t. The mean profile is denoted by $\widehat{\vec{\mu}}$ and to describe the variability we form the *anomalies*

$$\vec{\mathbf{x}}'_t = \vec{\mathbf{x}}_t - \widehat{\vec{\mu}}.$$

These anomalies are then *expanded* into a finite series

$$\vec{\mathbf{x}}'_t = \sum_{i=1}^{k} \widehat{\alpha}_{i,t} \widehat{\vec{e}}^{\,i} \tag{13.1}$$

with time coefficients $\widehat{\alpha}_{i,t}$ and fixed *patterns* $\widehat{\vec{e}}^{\,i}$. Equality is usually only possible when $k = m$, but the variance of the time coefficients $\widehat{\alpha}_{i,t}$ usually decreases quickly with increasing index i, so that good approximations are usually possible for k much less than m. The patterns are chosen to be *orthogonal* so that optimal coefficients $\widehat{\alpha}_{i,t}$ are obtained by simply projecting the anomalies $\vec{\mathbf{x}}'_t$ onto the patterns $\widehat{\vec{e}}^{\,i}$. Moreover, the patterns can be specified such that the *error*

$$\sum_t \left(\vec{\mathbf{x}}'_t - \sum_{i=1}^{k} \widehat{\alpha}_{i,t} \widehat{\vec{e}}^{\,i} \right)^2$$

is minimal. The lag-0 sample cross-correlations of the optimal time coefficients are all zero,

$$\sum_t \widehat{\alpha}_{i,t} \widehat{\alpha}_{j,t} = 0$$

for $i \neq j$. The patterns $\widehat{\vec{e}}^{\,j}$ are estimated *Empirical Orthogonal Functions*.[3] The coefficients $\widehat{\alpha}_i$ are the *EOF coefficients*.[4]

[1] A misnomer.

[2] The mathematics in this subsection are explained in more detail in Sections 13.1 and 13.2.

[3] Note that the 'functions' $\widehat{\vec{e}}^{\,k}$ are really *vectors* and not *functions*.

[4] Statisticians refer to the EOF coefficients as *principal components*.

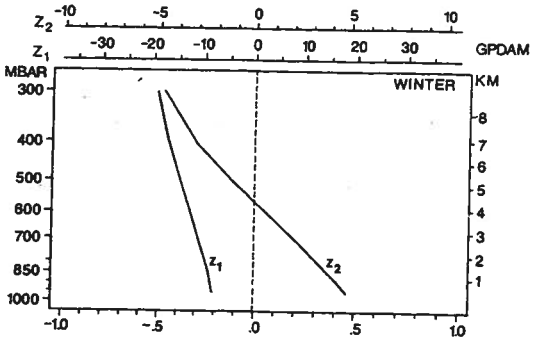

Figure 13.1: *The first two EOFs, labelled z_1 and z_2, of the daily geopotential height over Berlin in winter. From Fraedrich and Dümmel [125].*

The analysis of daily Berlin radiosonde data showed that only two patterns are required to describe most of the variability in the observed geopotential height profiles in winter (NDJF) as well as in summer (MJJA). In winter, the first EOF represents 91.2% of the variance (92.6% in summer), and the second EOF represents an additional 8.2% of the variance (7% in summer).[5] The remaining seven EOFs, which together with the first two EOFs span the full nine-dimensional space, represent only 0.6% of the variance of the height profiles (0.4% in summer). Thus, only two coefficient time series are required to represent the essential information in the time series of geopotential height at the nine levels. Instead of dealing with 1080 numbers per season, only $2 \times 120 = 240$ are needed. This demonstrates one of the advantages of EOFs, namely the ability to often identify a small subspace that contains most of the dynamics of the observed system.[6]

Another advantage is that the patterns can sometimes be seen as *modes of variability*. In the present example the two patterns $\vec{e}^{\,1}$ and $\vec{e}^{\,2}$ may be identified with the equivalent barotropic mode and the first baroclinic mode of the tropospheric circulation: The first patterns in winter (Figure 13.1) as well as in summer have the same sign throughout the troposphere, that is, they exhibit an equivalent barotropic structure. The second EOF, however, changes sign in the middle of the troposphere: it represents the first baroclinic mode.[7]

13.0.2 'Complex' EOFs. EOFs may be derived from real- or complex-valued random vectors. The latter results in complex-valued EOFs. The 'Complex EOF Analysis' (CEOF) described in the climate literature (see [181]) is a special case of the EOF analysis of complex random vectors. The time order of the observations is important for these 'CEOFs,' or 'Frequency Domain EOFs,' since they are the EOFs of a complexified time series. In contrast, the time order of observations is irrelevant in ordinary complex EOF analysis. The original real-valued time series is made complex by adding its Hilbert transform (see Section 16.2) as the imaginary component. The Hilbert transform can be thought of as the time derivative of the original process so that the EOF analysis of the complexified process reveals properties of the variability of the state and its change at the same time. To avoid confusion with the ordinary complex EOF analysis we refer to these EOFs as *Hilbert EOFs* (see Section 16.3).

13.1 Definition of Empirical Orthogonal Functions

13.1.1 Overview. EOFs are introduced formally in this section as parameters of the distribution of an m-dimensional random vector $\vec{\mathbf{X}}$.[8] For the sake of brevity we assume $\vec{\mu} = 0$. We first construct the first EOF, which is the most powerful single pattern in representing the variance of $\vec{\mathbf{X}}$. The idea is easily generalized to several patterns and in [13.1.3] the calculations are condensed into a theorem.

13.1.2 The First EOF. The first step is to find one 'pattern' $\vec{e}^{\,1}$, with $\|\vec{e}^{\,1}\| = 1$, such that

$$\epsilon_1 = \mathcal{E}\left(\|\vec{\mathbf{X}} - \langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle \vec{e}^{\,1}\|^2\right) \tag{13.2}$$

is minimized.[9] Equation (13.2) describes the projection of the random vector $\vec{\mathbf{X}}$ onto a one-dimensional subspace spanned by the fixed vector

[5] When we say that an expansion Y 'represents' p% of the variance of X, we mean that the variance of $Y-X$ is $(100-p)$% of the variance of X. The word 'explains' is often used instead of the word 'represents' in the literature. This is misleading since nothing is explained causally; only part of the variability of X has been described by Y.

[6] The assumption that the subspace with maximum variance coincides with the dynamically active subspace is arbitrary. In general, it will not be valid and counter examples can easily be constructed. However, in climate research, it is often reasonable to make this assumption. An example demonstrating the dynamical dominance of EOFs is given by Selten [343].

[7] A similar result for the vertical structure of the shelf ocean has been reported by Kundu, Allen, and Smith [234].

[8] Mainly based on [392].

[9] $\|\cdot\|$ denotes the vector norm, and $\langle\cdot,\cdot\rangle$ denotes the inner product. Note that $\|\vec{\mathbf{X}}\| = \langle\vec{\mathbf{X}}, \vec{\mathbf{X}}\rangle$. See Appendix B.

$\vec{e}^{\,1}$. Minimizing ϵ_1 is equivalent to the maximizing of the variance of $\vec{\mathbf{X}}$ that is contained in this subspace:

$$\begin{aligned}\epsilon_1 &= \mathcal{E}\left(\|\vec{\mathbf{X}}\|^2 - 2\langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle^*\vec{\mathbf{X}}^\dagger\vec{e}^{\,1} \right.\\ &\qquad \left. + \langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle^*\langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle\right)\\ &= \mathcal{E}\left(\|\vec{\mathbf{X}}\|^2 - \langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle^*\langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle\right)\\ &= \mathrm{Var}(\vec{\mathbf{X}}) - \mathrm{Var}\left(\langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle\right),\end{aligned}$$

where the variance of the random vector $\vec{\mathbf{X}}$ is defined to be the sum of variances of the elements of $\vec{\mathbf{X}}$.[10] Note that

$$\mathrm{Var}\left(\langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle\right) = \vec{e}^{\,1\dagger}\mathbf{\Sigma}\vec{e}^{\,1},$$

where $\mathbf{\Sigma}$ is the covariance matrix of $\vec{\mathbf{X}}$. Then minimization of equation (13.2), under the constraint $\|\vec{e}^{\,1}\| = 1$, leads to

$$\begin{aligned}&\frac{d}{d\vec{e}^{\,1}}\left[-\vec{e}^{\,1\dagger}\mathbf{\Sigma}\vec{e}^{\,1} + \lambda\left(\vec{e}^{\,1\dagger}\vec{e}^{\,1} - 1\right)\right]\\ &\quad = \ 2\mathbf{\Sigma}\vec{e}^{\,1} + 2\lambda\vec{e}^{\,1} = 0\end{aligned}$$

where λ is the Lagrange multiplier associated with the constraint $\|\vec{e}^{\,1}\| = 1$.[11] Thus, $\vec{e}^{\,1}$ is an eigenvector with a corresponding eigenvalue λ, of the covariance matrix $\mathbf{\Sigma}$. But $\mathbf{\Sigma}$ has m eigenvectors. Therefore, to minimize ϵ_1, we select the eigenvector that maximizes

$$\begin{aligned}\mathrm{Var}\left(\langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle\right) &= \vec{e}^{\,1\dagger}\mathbf{\Sigma}\vec{e}^{\,1}\\ &= \vec{e}^{\,1\dagger}\lambda\vec{e}^{\,1} = \lambda.\end{aligned}$$

Thus ϵ_1 is minimized when $\vec{e}^{\,1}$ is an eigenvector of $\mathbf{\Sigma}$ associated with its largest eigenvalue λ.[12] This 'pattern' is the first EOF.[13]

13.1.3 More EOFs. Having found the first EOF, we now repeat the exercise by finding the 'pattern' $\vec{e}^{\,2}$ that minimizes

$$\epsilon_2 = \mathcal{E}\left(\|(\vec{\mathbf{X}} - \langle\vec{\mathbf{X}}, \vec{e}^{\,1}\rangle\vec{e}^{\,1}) - \langle\vec{\mathbf{X}}, \vec{e}^{\,2}\rangle\vec{e}^{\,2}\|^2\right)$$

subject to the constraint that $\|\vec{e}^{\,2}\| = 1$. The result is that $\vec{e}^{\,2}$ is the eigenvector of $\mathbf{\Sigma}$ that corresponds to its second largest eigenvalue λ_2.[14] This second pattern is orthogonal to the first because the eigenvectors of a Hermitian matrix are orthogonal to one another.

13.1.4 Theorem. The following theorem results from the analysis presented so far.

Let $\vec{\mathbf{X}}$ be an m-dimensional random vector with mean $\vec{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ be the eigenvalues of $\mathbf{\Sigma}$ and let $\vec{e}^{\,1}, \ldots, \vec{e}^{\,m}$ be the corresponding eigenvectors of unit length. Since $\mathbf{\Sigma}$ is Hermitian, the eigenvalues are non-negative and the eigenvectors are orthogonal.

(i) *The k eigenvectors that correspond to $\lambda_1, \ldots, \lambda_k$ minimize*

$$\epsilon_k = \mathcal{E}\left(\|(\vec{\mathbf{X}} - \vec{\mu}) - \textstyle\sum_{i=1}^{k}\langle\vec{\mathbf{X}} - \vec{\mu}, \vec{e}^{\,i}\rangle\vec{e}^{\,i}\|^2\right). \tag{13.3}$$

$$\textbf{(ii)} \quad \epsilon_k = \mathrm{Var}(\vec{\mathbf{X}}) - \sum_{i=1}^{k}\lambda_i. \tag{13.4}$$

$$\textbf{(iii)} \quad \mathrm{Var}(\vec{\mathbf{X}}) = \sum_{i=1}^{m}\lambda_i. \tag{13.5}$$

The total variance of $\vec{\mathbf{X}}$ is broken up into m components. Each of these components is obtained by projecting $\vec{\mathbf{X}}$ onto one of the EOFs $\vec{e}^{\,i}$. The variance contribution of the kth component to the total variance $\sum_j \lambda_j$ is just λ_k. In relative terms, the proportion of the total variance represented by EOF k is $\lambda_k / \sum_j \lambda_j$. This proportion may be given as a percentage.

If the components are ordered by the size of the eigenvalues then the first component is the most important in representing variance, the second is the second most important and so forth.

Equation (13.3) gives the mean squared error ϵ_k that is incurred when approximating the full m-dimensional random vector $\vec{\mathbf{X}}$ in a k-dimensional subspace spanned by the first k EOFs. The construction of the EOFs ensures that the approximation is optimal; the use of any other k-dimensional subspace will lead to mean squared errors at least as large as ϵ_k.

13.1.5 Properties of the EOF Coefficients. The EOF coefficients, or *principal components*,

$$\alpha_i = \langle\vec{\mathbf{X}}, \vec{e}^{\,i}\rangle = \vec{\mathbf{X}}^{\mathrm{T}}\vec{e}^{\,i*} = \vec{e}^{\,i\dagger}\vec{\mathbf{X}} \tag{13.6}$$

[10] That is, if $\vec{\mathbf{X}}$ has covariance matrix $\mathbf{\Sigma}$, then we define $\mathrm{Var}\left(\vec{\mathbf{X}}\right) = \mathrm{tr}(\mathbf{\Sigma})$.

[11] Graybill [148, Section 10.8], describes the differentiation of quadratic forms.

[12] Recall (see Appendix B) that all eigenvalues of the Hermitian matrix $\mathbf{\Sigma} = \mathcal{E}(\vec{\mathbf{X}}\vec{\mathbf{X}}^\dagger)$ are real and non-negative.

[13] The pattern is unique up to sign if $\mathbf{\Sigma}$ has only one eigenvector that corresponds to eigenvalue λ. Otherwise, the pattern can be any vector with unit norm that is spanned by the eigenvectors corresponding to λ. In this case, the EOF is said to be *degenerate*. See Appendix B.

[14] Note that $\lambda_1 = \lambda_2$ if $\vec{e}^{\,1}$ is degenerate. In fact, if λ_1 has k linearly independent eigenvectors, then k of the m eigenvalues of $\mathbf{\Sigma}$ will be equal to λ.

are uncorrelated, and hence independent when $\vec{\mathbf{X}}$ is multivariate normal. In fact, for $i \neq j$,

$$\begin{aligned}\mathrm{Cov}(\alpha_i, \alpha_j) &= \mathcal{E}\big(\langle(\vec{\mathbf{X}} - \vec{\mu}), \vec{e}^{\,i}\rangle\langle(\vec{\mathbf{X}} - \vec{\mu}), \vec{e}^{\,j}\rangle^*\big) \\ &= \vec{e}^{\,i\dagger}\mathcal{E}\big((\vec{\mathbf{X}} - \vec{\mu})(\vec{\mathbf{X}} - \vec{\mu})^\dagger\big)\vec{e}^{\,j} \\ &= \vec{e}^{\,i\dagger}\Sigma\vec{e}^{\,j} \\ &= \lambda_j \vec{e}^{\,i\dagger}\vec{e}^{\,j} = 0\end{aligned}$$

Therefore, the variance of $\mathbf{X}_k$, the kth component of $\vec{\mathbf{X}}$, can also be decomposed into contribution from the individual EOFs as

$$\mathrm{Var}(\mathbf{X}_k) = \sum_{i=1}^{m} \lambda_i |e_k^i|^2. \tag{13.7}$$

If the elements of $\vec{\mathbf{X}}$ represent locations in space, the spatial distribution of variance can be visualized by plotting $\mathrm{Var}(\mathbf{X}_k)$ as a function of location. Similarly, the variance contribution from the ith EOF can be visualized by plotting $\lambda_i |e_k^i|^2$ or $\lambda_i |e_k^i|^2/\mathrm{Var}(\mathbf{X}_k)$ as a function of location.

13.1.6 Interpretation. The bulk of the variance of $\vec{\mathbf{X}}$ can often be represented by the first few EOFs. If the original variable has m components the approximation of $\vec{\mathbf{X}}$ by $\vec{\alpha} = (\alpha_1, \ldots, \alpha_k)$, with $k \ll m$, leads to a significant reduction of the amount of data while retaining most of the variance. It was shown in the introductory example of Berlin geopotential height [13.0.2] that just two EOFs represent almost all of the information in the data set.

The physical interpretation of EOFs is limited by a fundamental constraint. While it is often possible to clearly associate the first EOF with a known physical process, this is much more difficult with the second (and higher-order) EOF because it is constrained to be orthogonal to the first EOF. However, real-world processes do not need to have orthogonal patterns or uncorrelated indices. In fact, the patterns that most efficiently represent variance do not necessarily have anything to do with the underlying dynamical structure.

13.1.7 Vector Notation. The random vector $\vec{\mathbf{X}}$ may conveniently be written in vector notation by

$$\vec{\mathbf{X}} = \mathcal{P}\vec{\alpha} \tag{13.8}$$

where $\mathcal{P}$ is the $m \times m$ matrix

$$\mathcal{P} = \big(\vec{e}^{\,1}|\vec{e}^{\,2}|\cdots|\vec{e}^{\,m}\big) \tag{13.9}$$

that has EOFs in its columns, and $\vec{\alpha}$ is the m-dimensional (column) vector of EOF coefficients $\alpha_1, \ldots, \alpha_m$. Because the EOFs are orthonormal, the expression (13.8) may be inverted to obtain

$$\vec{\alpha} = \mathcal{P}^\dagger\vec{\mathbf{X}}, \tag{13.10}$$

where $\mathcal{P}^\dagger$ is the conjugate transpose of $\mathcal{P}$. Another consequence of the orthonormality of the EOFs is that

$$\begin{aligned}\Sigma &= \mathrm{Cov}(\vec{\mathbf{X}}, \vec{\mathbf{X}}) \\ &= \mathcal{P}\mathrm{Cov}(\vec{\alpha}, \vec{\alpha})\mathcal{P}^\dagger \\ &= \mathcal{P}\Lambda\mathcal{P}^\dagger\end{aligned}$$

where Λ is the diagonal $m \times m$ matrix composed of the eigenvalues of Σ,

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m).$$

It therefore follows that

$$\begin{aligned}\mathrm{Var}(\vec{\mathbf{X}}) &= \sum_{k=1}^{m} \mathrm{Var}(\mathbf{X}_k) \\ &= \mathrm{tr}(\Sigma) \\ &= \mathrm{tr}(\mathcal{P}\Lambda\mathcal{P}^\dagger) \\ &= \mathrm{tr}(\Lambda) = \sum_{k=1}^{m} \lambda_k.\end{aligned}$$

It also follows that the eigenvalues are the m roots of the mth degree *characteristic polynomial* $p_\Sigma(\lambda) = \det(\Sigma - \lambda\mathcal{I})$, where $\mathcal{I}$ is the $m \times m$ identity matrix. In fact

$$\begin{aligned}p_\Sigma(\lambda) &= \det(\mathcal{P}\Lambda\mathcal{P}^\dagger - \lambda\mathcal{P}\mathcal{P}^\dagger) \\ &= \det\big(\mathcal{P}(\Lambda - \lambda\mathcal{I})\mathcal{P}^\dagger\big) \\ &= \det(\Lambda - \lambda\mathcal{I}) = \prod_{i=1}^{m}(\lambda_i - \lambda).\end{aligned} \tag{13.11}$$

13.1.8 Degeneracy. As noted above, EOFs are not uniquely determined. If λ_o is a root of multiplicity 1 of $p_\Sigma(\lambda)$ and $\vec{e}$ is a corresponding (normalized) eigenvector, then $\vec{e}$ is unique up to sign, and either $\vec{e}$ or $-\vec{e}$ is chosen as the EOF that corresponds to λ_o. On the other hand, if λ_o is a root of multiplicity k, the solution space of

$$\Sigma\vec{e} = \lambda_o\vec{e}$$

is of dimension k. The solution space is uniquely determined in the sense that it is orthogonal to the space spanned by the $m - k$ eigenvectors of Σ that correspond to eigenvalues $\lambda_i \neq \lambda_o$. But any orthonormal basis $\vec{e}^{\,1}, \ldots, \vec{e}^{\,k}$ for the solution space can be used as EOFs. In this case the EOFs are said to be *degenerate*. (An example is discussed in [13.1.9].)

Degeneracy can either be bad or good news. It is bad news if the EOFs are estimated from a sample

of iid realizations of $\vec{\mathbf{X}}$. Then degeneracy is mostly a nuisance, because the patterns, which may represent independent processes in the underlying dynamics, can not be disentangled.

However, degeneracy may be good news if the EOFs are estimates from a realization of a stochastic process $\vec{\mathbf{X}}_t$. Suppose, for example, that $p_\Sigma(\lambda)$ has a root of multiplicity 2. By construction, the cross-correlation of the two corresponding EOF coefficient time series will be zero at lag-0. But this does not imply that the lagged cross-correlations will be zero, and, in fact, they are often nonzero. This means that a pair of EOFs and their coefficient series could represent a signal that is propagating in space.

The representation of such a spatially propagating signal requires two patterns whose coefficients vary coherently and are 90° out-of-phase. The two patterns representing a propagating signal are not uniquely determined; indeed if any two patterns represent the signal, then any linear combination of the two do so as well. Therefore, degeneracy is a necessary condition for the description of such signals.

13.1.9 Examples. To demonstrate the mathematics of EOF analysis and the phenomenon of degeneracy we now consider the case of a random vector

$$\vec{\mathbf{X}} = \sum_{k=1}^{m} \alpha_k \vec{p}^{\,k} \tag{13.12}$$

where coefficients α_k are uncorrelated real univariate random variables and $\vec{p}^{\,1}, \ldots, \vec{p}^{\,m}$ are fixed orthonormal vectors. For simplicity we assume that the αs have mean zero. Then the covariance matrix of $\vec{\mathbf{X}}$ is

$$\begin{aligned} \mathbf{\Sigma} &= \mathcal{E}\Big(\Big(\textstyle\sum_k \alpha_k \vec{p}^{\,k}\Big)\Big(\sum_l \alpha_l \vec{p}^{\,l}\Big)^{\mathrm{T}}\Big) \\ &= \sum_j \mathrm{Var}(\alpha_j)\vec{p}^{\,j}\vec{p}^{\,j^{\mathrm{T}}}. \end{aligned} \tag{13.13}$$

It is easily verified that $\mathrm{Var}(\alpha_k)$ is an eigenvalue of this covariance matrix with eigenvector $\vec{p}^{\,k}$:

$$\Big(\sum_j \mathrm{Var}(\alpha_j)\vec{p}^{\,j}\vec{p}^{\,j^{\mathrm{T}}}\Big)\vec{p}^{\,k} = \mathrm{Var}(\alpha_k)\vec{p}^{\,k}.$$

Thus, the chosen orthonormal vectors are the EOFs of the random vector (13.12). The ordering is determined by the variance of the uncorrelated univariate random variables α_k.

The example has two merits. First it may be used as a recipe for constructing random vectors with a given EOF structure. To do so one has to select a set of orthonormal vectors and associate them with a set of uncorrelated random variables.

The example may also be used to demonstrate the phenomenon of degeneracy. To do so, we assume that all αs have variance 1. Then, the EOFs are degenerate and may be replaced by any other set of orthonormal vectors. One such set of orthonormal vectors are the unit vectors $\vec{u}^{\,k}$ with a 1 in the kth row and zeros elsewhere. Then, the representation (13.8), with $\mathcal{P} = (\vec{p}^{\,1}|\cdots|\vec{p}^{\,m})$, is transformed as

$$\vec{\mathbf{X}} = \mathcal{P}\vec{\alpha} = \Big(\mathcal{P}\mathcal{P}^{\mathrm{T}}\Big)(\mathcal{P}\vec{\alpha}) = \mathcal{U}\vec{\beta}, \tag{13.14}$$

where the new EOFs are the columns of

$$\mathcal{U} = \mathcal{P}\mathcal{P}^{\mathrm{T}} = \mathcal{I} = (\vec{u}^{\,1}|\cdots|\vec{u}^{\,m})$$

and the EOF coefficients are given by

$$\vec{\beta} = \mathcal{P}\vec{\alpha}.$$

These coefficients are uncorrelated as well because of $\mathrm{Var}(\alpha_k) = 1$ for all k:

$$\begin{aligned} \mathrm{Cov}(\vec{\beta},\vec{\beta}) &= \mathrm{Cov}(\mathcal{P}\vec{\alpha},\mathcal{P}\vec{\alpha}) \\ &= \mathcal{P}\,\mathrm{Cov}(\alpha,\alpha)\mathcal{P}^{\mathrm{T}} \\ &= \mathcal{P}\mathcal{I}\mathcal{P}^{\mathrm{T}} = \mathcal{I}. \end{aligned}$$

Obviously, the only meaningful information the EOF analysis offers in this case is that there is no preferred direction in the phase space. The only property that matters is the uniformity of the variance in all directions.

13.1.10 Coordinate Transformations. Let us consider two m-variate random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Z}}$ that are related to each other by

$$\vec{\mathbf{Z}} = \mathcal{L}\vec{\mathbf{X}} \tag{13.15}$$

where $\mathcal{L}$ is an invertible matrix so that $\vec{\mathbf{X}} = \mathcal{L}^{-1}\vec{\mathbf{Z}}$. Both vectors represent the same information but the data are given in different coordinates. The covariance matrix of $\vec{\mathbf{Z}}$ is

$$\mathbf{\Sigma}_{ZZ} = \mathcal{L}\mathbf{\Sigma}_{XX}\mathcal{L}^{\dagger}, \tag{13.16}$$

for $\mathcal{L}^{\dagger}$ the conjugate transpose of $\mathcal{L}$. Suppose the *transformation* is *orthogonal* (i.e., $\mathcal{L}^{-1} = \mathcal{L}^{\dagger}$), and also let λ be an eigenvalue of $\mathbf{\Sigma}_{XX}$ and let $\vec{e}^{\,X}$ be the corresponding eigenvector. Then, since $\mathbf{\Sigma}_{XX}\vec{e}^{\,X} = \lambda\vec{e}^{\,X}$,

$$\begin{aligned} \mathbf{\Sigma}_{ZZ}\mathcal{L}\vec{e}^{\,X} &= \mathcal{L}\mathbf{\Sigma}_{XX}\mathcal{L}^{\dagger}\mathcal{L}\vec{e}^{\,X} \\ &= \mathcal{L}\mathbf{\Sigma}_{XX}\vec{e}^{\,X} \\ &= \lambda\mathcal{L}\vec{e}^{\,X}. \end{aligned}$$

Thus λ is also an eigenvalue of Σ_{ZZ} and the EOFs of $\vec{\mathbf{Z}}$ are related to those of $\vec{\mathbf{X}}$ through

$$\vec{e}^{\,Z} = \mathcal{L}\vec{e}^{\,X}. \tag{13.17}$$

Thus eigenvectors are transformed (13.17) just as a random vector is transformed (13.15).

Another consequence of using an orthogonal transformation is that the EOF coefficients are invariant. To see this, let $\mathcal{P}_X$ be the matrix composed of the $\vec{\mathbf{X}}$-EOFs $\vec{e}^{\,X}$ and let $\mathcal{P}_Z$ the the corresponding matrix of $\vec{\mathbf{Z}}$-EOFs. We see from equation (13.17) that

$$\mathcal{P}_Z = \mathcal{L}\mathcal{P}_X. \tag{13.18}$$

Using equation (13.10) and transformation (13.15), the vector of $\vec{\mathbf{X}}$-EOF coefficients

$$\begin{aligned}\vec{\alpha}_X &= \mathcal{P}_X^\dagger \vec{\mathbf{X}} = \mathcal{P}_X^\dagger \mathcal{L}^\dagger \vec{\mathbf{Z}} = (\mathcal{L}\mathcal{P})^\dagger \vec{\mathbf{Z}} = \mathcal{P}_Y^\dagger \vec{\mathbf{Z}} \\ &= \vec{\alpha}_Z\end{aligned}$$

is seen to be equal to the vector of $\vec{\mathbf{Z}}$-EOF coefficients. Thus the EOF coefficients are invariant under *orthogonal transformations*. They are generally not invariant under non-orthogonal transformations. This becomes important when different variables such as precipitation and temperature are combined in a data vector. The EOFs of such a random vector depend on the units in which the variables are expressed.

A special case of transformation (13.15) occurs when $\vec{\mathbf{X}}$ has already been transformed into EOF coordinates using (13.10),

$$\vec{\mathbf{Z}} = \vec{\alpha} = \mathcal{P}^\dagger \vec{\mathbf{X}}.$$

That is, $\mathcal{L} = \mathcal{P}^\dagger$. Using transformation (13.18), we see that the $\vec{\alpha}$-EOFs are

$$\mathcal{P}_\alpha = \mathcal{L}\mathcal{P} = \mathcal{P}^\dagger\mathcal{P} = \mathcal{I}.$$

Thus, in the new coordinates the EOFs are unit vectors. This fact may be used to test EOF programs.

13.1.11 Further Aspects. Some other aspects of EOFs are worth mentioning.

- Empirical Orthogonal Functions may be generalized to continuous functions, in which case they are known as *Karhunen-Loève* functions. The standard inner product $\langle\cdot,\cdot\rangle$ is replaced by an integral, and the eigenvalue problem is no longer a matrix problem but an operator problem. (See, e.g., North et al. [296].)

- The EOFs of some random vectors or random functions are given by sets of analytic orthogonal functions. For instance, if the covariance structure of a spatial process is independent of the location, then the EOFs on a continuous or regularly discretized sphere (circle) are the spherical harmonics (trigonometric functions). See North et al. [296].

- The analysed vector $\vec{\mathbf{X}}$ may be a combination of small vectors that are expressed on different scales, such as temperature and precipitation or geopotential height at 700 and 200 hPa. Then the technique is sometimes called *Combined Principal Component Analysis* (see, e.g., Bretherton, Smith, and Wallace [64]). Vector $\vec{\mathbf{X}}$ might also consist of smaller vectors representing a single field observed at different times, in which case the technique is called *Extended EOF Analysis* (EEOF; see Weare and Nasstrom [417]) or *Multichannel Singular Spectrum Analysis* (MSSA; see Section 13.6).

- Any m-dimensional vector $\vec{\mathbf{y}}$ can be projected onto an EOF $\vec{e}$ of $\vec{\mathbf{X}}$ by computing the inner product $\langle\vec{\mathbf{y}}, \vec{e}\,\rangle$. Vector $\vec{\mathbf{y}}$ can then by approximated by $\vec{\mathbf{y}} \approx \sum_{i=1}^{k} \langle\vec{\mathbf{y}}, \vec{e}^{\,i}\rangle \vec{e}^{\,i}$.

- Where are the units? When we expand the random vector $\vec{\mathbf{X}}$ into EOFs

$$\vec{\mathbf{X}} \approx \sum_{i=1}^{k} \alpha_i \vec{e}^{\,i} \tag{13.19}$$

with $\alpha_i = \langle\vec{\mathbf{X}}, \vec{e}^{\,i}\rangle$, where do we place the units of $\vec{\mathbf{X}}$ on the right side of approximation (13.19)? Formally the answer is that the coefficients carry the units while the patterns are dimensionless. However, in practice approximation (13.19) is often replaced by

$$\vec{\mathbf{X}} \approx \sum_{i=1}^{k} \alpha_i^+ \vec{e}^{\,i+} \tag{13.20}$$

with re-normalized coefficients

$$\alpha_i^+ = \frac{1}{\sqrt{\lambda_i}}\alpha_i \tag{13.21}$$

and patterns

$$\vec{e}^{\,i+} = \sqrt{\lambda_i}\,\vec{e}^{\,i} \tag{13.22}$$

so that $\mathrm{Var}(\alpha_i^+) = 1$. The re-normalized pattern then carries the units of $\vec{\mathbf{X}}$, and represents a 'typical' anomaly pattern if we regard $\alpha_i^+ = \pm 1$ as a 'typical event'.

The decomposition of the local variance, as given by equation (13.7), takes a particularly simple form with this normalization, namely

$$\mathrm{Var}(\mathbf{X}_k) = \sum_{i=1}^{m} |e_k^{i+}|^2. \tag{13.23}$$

Note that the coefficient α_i^+ can be expressed as

$$\alpha_i^+ = \frac{1}{\lambda_i} \langle \vec{\mathbf{X}}, \vec{e}^{\,i+} \rangle. \tag{13.24}$$

13.2 Estimation of Empirical Orthogonal Functions

13.2.1 Outline. After having defined the eigenvalues and EOFs of random vector $\vec{\mathbf{X}}$ as parameters that characterize its covariance matrix, the question naturally arises as to how to *estimate* these parameters from sample $\{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n\}$ of realizations of $\vec{\mathbf{X}}$. It turns out that useful estimators may be defined by replacing the covariance matrix Σ with the sample covariance matrix $\widehat{\Sigma}$ and by replacing the expectation operator $\mathcal{E}(\cdot)$ with averaging over the sample. An important little trick for reducing the amount of calculation when the sample size n is less than the dimension of $\vec{\mathbf{X}}$ (as is often true) is presented in [13.2.5]. A computational alternative to solving the eigenproblem is to perform a singular value decomposition [13.2.8].

13.2.2 Strategies for Estimating EOFs. The eigenvalues and EOFs are parameters that characterize the covariance matrix of a random vector $\vec{\mathbf{X}}$. In practice, the distribution of $\vec{\mathbf{X}}$, and thus the covariance matrix Σ and its eigenvalues and eigenvectors, is unknown. They must therefore be estimated from a finite sample $\{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n\}$.

There are two reasonable approaches for estimation.

- Since the eigenvalues and EOFs characterize the covariance matrix Σ of $\vec{\mathbf{X}}$, one reasonable approach is to estimate the covariance matrix and then estimate the eigenvalues λ_i and eigenvectors $\vec{e}^{\,i}$ with the eigenvalues λ_j and the eigenvectors $\vec{e}^{\,j}$ of the estimated covariance matrix $\widehat{\Sigma}$.
- After [13.1.3] the EOFs form an orthonormal set of vectors that is most efficient in representing the variance of $\vec{\mathbf{X}}$ (13.3). Thus another reasonable approach is to use a set of orthonormal vectors that represent as much of the sample variance of the finite sample as possible.

The two approaches are equivalent and lead to the following.

13.2.3 Theorem. *Let* $\widehat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\vec{\mathbf{x}}_j - \widehat{\mu})(\vec{\mathbf{x}}_j - \widehat{\mu})^\dagger$, *where* $\dagger$ *indicates the conjugate transpose and* $\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n}\vec{\mathbf{x}}_j$, *derived from a sample* $\{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n\}$ *be the estimated covariance matrix of n realizations of* $\vec{\mathbf{X}}$. *Let* $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_m$ *be the eigenvalues of* $\widehat{\Sigma}$ *and let* $\widehat{\vec{e}}^{\,1}, \ldots, \widehat{\vec{e}}^{\,m}$ *be corresponding eigenvectors of unit length. Since* $\widehat{\Sigma}$ *is Hermitian, the eigenvalues are non-negative and the eigenvectors are orthogonal.*

(i) *The k eigenvectors that correspond to* $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_k$ *minimize*

$$\widehat{\epsilon}_k = \sum_{j=1}^{n} \left| \vec{\mathbf{x}}_j - \sum_{i=1}^{k} \langle \vec{\mathbf{x}}_j, \widehat{\vec{e}}^{\,i} \rangle \widehat{\vec{e}}^{\,i} \right|^2. \tag{13.25}$$

$$\text{(ii)} \quad \widehat{\epsilon}_k = \widehat{\mathrm{Var}}(\vec{\mathbf{X}}) - \sum_{j=1}^{k} \widehat{\lambda}_j. \tag{13.26}$$

$$\text{(iii)} \quad \widehat{\mathrm{Var}}(\vec{\mathbf{X}}) = \sum_{j=1}^{m} \widehat{\lambda}_j, \tag{13.27}$$

where $\widehat{\mathrm{Var}}(\vec{\mathbf{X}}) = \mathrm{tr}(\widehat{\Sigma})$.

13.2.4 The Estimated Covariance Matrix $\widehat{\Sigma}$. The covariance between the jth and kth elements of $\vec{\mathbf{X}}$ is estimated by

$$\widehat{\sigma}_{jk} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{ji} - \overline{\mathbf{x}}_j)(\mathbf{x}_{ki} - \overline{\mathbf{x}}_k),$$

where $\mathbf{x}_{ji}$ and $\mathbf{x}_{ki}$ are the jth and kth elements of $\vec{\mathbf{x}}_i$. This sum of products can be expressed as a *quadratic form*:[15]

$$\widehat{\Sigma} = \frac{1}{n}\mathcal{X}(\mathcal{I} - \frac{1}{n}\mathcal{J})(\mathcal{I} - \frac{1}{n}\mathcal{J})\mathcal{X}^\dagger \tag{13.28}$$

where $\mathcal{X}$ is the *data matrix*[16]

$$\mathcal{X} = \begin{pmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{m1} & \mathbf{x}_{m2} & \cdots & \mathbf{x}_{mn} \end{pmatrix}, \tag{13.29}$$

[15] A quadratic form is a matrix product of the form $\mathcal{A}\mathcal{A}^\dagger$ or $\mathcal{A}^\dagger\mathcal{A}$.

[16] Sometimes also called the design matrix.

and $\mathcal{X}^\dagger$ is the conjugate transpose of $\mathcal{X}$, $\mathcal{I}$ is the $n \times n$ identity matrix, and $\mathcal{J}$ is the $n \times n$ matrix composed entirely of units. The n columns of the $m \times n$ data matrix $\mathcal{X}$ are the sample vectors $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_n$; the rows mark the m coordinates in the original space. The matrix product $\mathcal{X}\mathcal{X}^\dagger$ is a square matrix even if $\mathcal{X}$ is not.

13.2.5 Theorem. The following theorem is often useful when computing eigenvalues and eigenvectors [391].

Let $\mathcal{A}$ be any $m \times n$ matrix. If λ is a nonzero eigenvalue of multiplicity s of $\mathcal{A}^\dagger\mathcal{A}$ with s linearly independent eigenvectors $\vec{e}^{\,1}, \ldots, \vec{e}^{\,s}$, then λ is also an s-fold eigenvalue of $\mathcal{A}\mathcal{A}^\dagger$ with s linearly independent eigenvectors $\mathcal{A}\vec{e}^{\,1}, \ldots, \mathcal{A}\vec{e}^{\,s}$.

A proof is given in Appendix M.

13.2.6 Recipe. The message of Theorem [13.2.5] is that the nonzero eigenvalues of $\mathcal{A}\mathcal{A}^\dagger$ are identical to those of $\mathcal{A}^\dagger\mathcal{A}$ and that the eigenvectors of the two matrices associated with nonzero eigenvalues are related through a simple linear relationship. Thus the following recipe may be used to estimate EOFs.

- If the sample size, n, is larger than the dimension of the problem, m, then the EOFs are calculated directly as the normalized eigenvectors of the $m \times m$ matrix $\frac{1}{n}\mathcal{X}(\mathcal{I} - \frac{1}{n}\mathcal{U})(\mathcal{I} - \frac{1}{n}\mathcal{U})\mathcal{X}^\dagger$.
- If the sample size, n, is smaller than the dimension of the problem, m, the EOFs may be obtained by first calculating the normalized eigenvectors $\vec{g}$ of the $n \times n$ matrix $\frac{1}{n}(\mathcal{I} - \frac{1}{n}\mathcal{J})\mathcal{X}^\dagger\mathcal{X}(\mathcal{I} - \frac{1}{n}\mathcal{J})$ and then computing the EOFs as

$$\vec{e} = \frac{\mathcal{X}(\mathcal{I} - \frac{1}{n}\mathcal{J})\vec{g}}{\|\mathcal{X}(\mathcal{I} - \frac{1}{n}\mathcal{J})\vec{g}\|}.$$

13.2.7 Properties of the Coefficients of the Estimated EOFs. There are several properties worth noting.

- As with the true EOFs, the estimated EOFs span the full m-dimensional vector space. Random vector $\vec{\mathbf{X}}$ can therefore be expanded in terms of the estimated EOFs as $\vec{\mathbf{X}} = \sum_{j=1}^{m} \widehat{\alpha}_j \vec{e}^{\,j}$, where

$$\widehat{\alpha}_j = \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,j} \rangle. \tag{13.30}$$

- When $\vec{\mathbf{X}}$ is multivariate normal, the distribution of $\widehat{\vec{\alpha}}$, where $\widehat{\vec{\alpha}}$ is the m-dimensional vector of EOF coefficients $\widehat{\alpha}_j$, conditional upon the samples used to estimate the EOFs is multivariate normal with mean $\mathcal{E}(\widehat{\vec{\alpha}}|\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m) = \widehat{\mathcal{P}}^\dagger\vec{\mu}$ and covariance matrix $\mathrm{Cov}(\widehat{\vec{\alpha}}, \widehat{\vec{\alpha}}|\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m) = \widehat{\mathcal{P}}^\dagger\Sigma\widehat{\mathcal{P}}$. Matrix $\widehat{\mathcal{P}}$, which has $\widehat{\vec{e}}^{\,j}$ in column j, is a complicated function of $\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m$.
- $\widehat{\lambda}_j$ is the variance of the EOF coefficients computed from the sample used to estimate the EOFs. That is, if $\widehat{\alpha}_{ji} = \langle \vec{\mathbf{X}}_i, \widehat{\vec{e}}^{\,j} \rangle$, then $\frac{1}{n}\sum_{i=1}^{n} |\widehat{\alpha}_{ji} - \overline{\widehat{\alpha}}_j|^2 = \widehat{\lambda}_j$.

  Note that $\widehat{\lambda}_j$ has at least two interpretations as a variance estimate. We could regard $\widehat{\lambda}_j$ as an estimate of the variance of the true EOF coefficient $\alpha_j = \langle \vec{\mathbf{X}}, \vec{e}^{\,j} \rangle$ (see [13.3.3]). Alternatively, we could view the estimated EOFs $\widehat{\vec{e}}^{\,j}$ as fixed, not quite optimal, proxies for $\vec{e}^{\,j}$. Then $\widehat{\lambda}_j$ could be viewed as an estimator of the variance of $\widehat{\alpha}_i = \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,j} \rangle$ when $\widehat{\vec{e}}^{\,j}$ is fixed (see [13.3.2]). These two variances are not equal, although they become asymptotically equivalent as $n \to \infty$. Thus, at least one of the interpretations makes $\widehat{\lambda}_j$ a biased estimator. In fact, they are both poor estimators when the sample is small. In the former case there is uncertainty because the EOFs must be estimated. In the latter case the EOFs are regarded as fixed, but there is a bias because independent data are not used to estimate $\mathrm{Var}(\widehat{\alpha}_i)$. See also [13.3.2,3].
- The sample covariance of a pair of EOF coefficients computed from the sample used to estimate the EOFs is zero. That is, $\frac{1}{n}\sum_{i=1}^{n}(\widehat{\alpha}_{ji} - \overline{\widehat{\alpha}}_j)(\widehat{\alpha}_{ki} - \overline{\widehat{\alpha}}_k)^* = 0$ if $j \neq k$.

  As with $\widehat{\lambda}_j$, the covariance has two interpretations. It correctly estimates the covariance of the true EOF coefficients $\alpha_j = \langle \vec{\mathbf{X}}, \vec{e}^{\,j} \rangle$ and $\alpha_k = \langle \vec{\mathbf{X}}, \vec{e}^{\,k} \rangle$. Alternatively, if we view the estimated EOFs $\widehat{\vec{e}}^{\,j}$ as being fixed, then it incorrectly estimates $\mathrm{Cov}(\widehat{\alpha}_j, \widehat{\alpha}_k)$. The latter, the (j, k) element of $\widehat{\mathcal{P}}^\dagger\Sigma\widehat{\mathcal{P}}$, can be substantially different from zero if $\widehat{\vec{e}}^{\,j}$ and $\widehat{\vec{e}}^{\,k}$ are computed from a small sample.

13.2.8 Gappy Data. Data are often incomplete, that is, there are irregularly distributed gaps in the data vectors caused by missing observations. Estimated EOFs and EOF coefficients can be derived in this case, but the procedure is slightly

different. Each element of Σ is estimated by forming sums of all available products

$$\widehat{\sigma}_{ij} = \frac{1}{|K_i \cap K_j|} \sum_{k \in K_i \cap K_j} (\mathbf{x}_{ki} - \widehat{\mu}_i)(\mathbf{x}_{kj} - \widehat{\mu}_j)^* \tag{13.31}$$

where $K_i = \{k\text{: component } i \text{ of } \vec{\mathbf{x}}_k \text{ is not missing}\}$, and where $\widehat{\mu}_i = \frac{1}{|K_i|} \sum_{k \in K_i} \mathbf{x}_{ki}$. The estimated EOFs are then the eigenvectors $\widehat{\vec{e}}^{\,i}$ of this covariance matrix estimate. The set $K_i \cap K_j$ is the set of all indices such that $\mathbf{x}_{ki}$ and $\mathbf{x}_{kj}$ are not missing. The $|\cdot|$ notation is used to indicate the size of the enclosed set.

The EOF coefficient $\widehat{\alpha}_i$ of a gappy data vector $\vec{\mathbf{x}}$ can not be obtained as a simple dot product of the gappy data vector $\vec{\mathbf{x}}$ and the estimated EOF $\widehat{\vec{e}}^{\,i}$, as in equation (13.30), but a least squares estimate can be obtained by choosing $\widehat{\alpha}_i$ to minimize $\|\vec{\mathbf{x}} - \widehat{\alpha}_i \widehat{\vec{e}}^{\,i}\|$. The least square estimate is given by

$$\widehat{\alpha}_i = \frac{\sum_{j \in K} \mathbf{x}_j \widehat{e}^{i*}_j}{\sum_{j \in K} |\widehat{e}^i_j|^2} \tag{13.32}$$

where $\mathbf{x}_j$ and $\widehat{e}^i_j$ are the jth components of $\mathbf{x}$ and $\vec{e}^{\,i}$, respectively, and where $K = \{j\text{: } \mathbf{x}_j \text{ is not missing}\}$. Note that equation (13.32) reduces to $\widehat{\alpha}_i = \langle \vec{\mathbf{x}}, \widehat{\vec{e}}^{\,i} \rangle$ when there are no gaps in $\mathbf{x}$.

13.2.9 Computing Eigenvalues and Eigenvectors. One approach to computing eigenvalues and eigenvectors is to use a 'canned' eigenanalysis routine such as those that are contained in EISPACK [352], IMSL [193], or NAG [298]. Press et al. [322, p. 454] discuss the origins of these routines and give further references.

An alternative approach uses Singular Value Decomposition (Appendix B, and see also Press et al. [322, pp. 51–63] and Kelly [218, 219]). The SVD of the conjugate transpose of the $m \times n$ *centred* data matrix

$$\mathcal{X}' = \mathcal{X}\left(\mathcal{I} - \frac{1}{n}\mathcal{J}\right), \tag{13.33}$$

where $\mathcal{X}$ is given by equation (13.29), $\mathcal{I}$ is the $n \times n$ identity matrix, and $\mathcal{J}$ is the $n \times n$ matrix of units, is

$$\mathcal{X}'^{\dagger} = \mathcal{U}\mathcal{S}\mathcal{V}^{\dagger}, \tag{13.34}$$

where $\mathcal{U}$ is $n \times m$, $\mathcal{S}$ and $\mathcal{V}$ are each $m \times m$, n is the sample size, and m is the dimension of $\vec{\mathbf{X}}$.[17] Since $n\widehat{\Sigma} = \mathcal{X}'\mathcal{X}'^{\dagger}$, we infer from equation (B.6) that the right singular vectors $\vec{v}^{\,i}$ are equal to the estimated EOFs $\widehat{\vec{e}}^{\,i}$. The singular values s_i are related to the estimated eigenvalues by $\widehat{\lambda}_i = \frac{1}{n}s_i^2$. The left singular vectors $\vec{u}^{\,i}$ are given by (B.5)

$$\vec{u}^{\,i} = \frac{1}{s_i}\mathcal{X}'^{\dagger}\vec{v}^{\,i}. \tag{13.35}$$

The kth column of $\mathcal{X}'$ represents the vector of deviations $\vec{\mathbf{x}}_k - \widehat{\vec{\mu}}$ so that

$$u^i_k = \frac{1}{s_i}(\vec{\mathbf{x}}_k - \widehat{\vec{\mu}})^{\dagger}\vec{v}^{\,i}. \tag{13.36}$$

Thus, u^i_k is the ith normalized EOF coefficient (13.21) of the anomalies $\vec{\mathbf{x}}_k - \widehat{\vec{\mu}}$. Note that the sample variance of the ith normalized EOF coefficient is

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{s_i}(\vec{\mathbf{x}}_k - \widehat{\vec{\mu}})^{\dagger}\widehat{\vec{e}}^{\,i}\right) &= \frac{1}{s_i^2}\widehat{\mathrm{Var}}\,(\widehat{\alpha}_i) \\ &= \frac{1}{s_i^2}\widehat{\lambda}_i = 1. \end{aligned} \tag{13.37}$$

Note also that equations (13.35)–(13.37) are only valid for those EOFs that correspond to nonzero eigenvalues. The number of nonzero eigenvalues, which is determined by the rank[18] of the centred data matrix, is no greater than $\min(m, n-1)$.

Thus SVD extracts the same information from the sample as a conventional EOF analysis.

13.3 Inference

13.3.1 General. We consider the reliability of eigenvalues and EOF estimates in this section. This is a somewhat different question from that which users generally have in mind when they enquire about the 'significance' of an EOF. The null hypothesis that is usually implicit in the latter is that the EOF in question describes only an aspect of the covariance structure of the 'noise' in the observed system, and the alternative hypothesis is that the EOF also describes at least part of the dynamics of the observed system. Unfortunately, discrimination between 'noise' and 'signal' in

[17] We have implicitly assumed here that $m \leq n$. The problem is approached similarly when $m > n$, except we begin by obtaining the SVD of $\mathcal{X}'$. Note also that in some texts $\mathcal{U}$ and $\mathcal{V}$ are $n \times n$ and $m \times m$ orthogonal matrices respectively and $\mathcal{S}$ is $n \times m$. The singular values are placed in the diagonal part of $\mathcal{S}$ and the rest of the matrix is zero. We use the decomposition given in (13.34) because it is commonly used in SVD subroutines (see, e.g., Press et al. [322]).

[18] The *rank* of a matrix is the dimension of the sub-space spanned by the columns of that matrix.

this way is fraught with difficulty. We discuss this further in [13.3.4]. However, we first briefly consider the variance of EOF coefficients in [13.3.2] and the bias of eigenvalue estimates in [13.3.3]. We consider the sampling error of the EOFs themselves in [13.3.5,6].

13.3.2 The Variance of EOF Coefficients of a Given Set of Estimated EOFs. Assume we are given a set of eigenvalues $\widehat{\lambda}_i$ and EOFs $\widehat{\vec{e}}^{\,i}$ that are derived from a finite sample $\{\vec{\mathbf{x}}_1 \dots \vec{\mathbf{x}}_n\}$. Then any random vector $\vec{\mathbf{X}}$ can be represented in the space spanned by these estimated EOFs by using the transformation $\widehat{\alpha} = \mathcal{P}^{\dagger}\vec{\mathbf{X}}$. The transformed random variables $\widehat{\alpha}_i = \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,i} \rangle$ have their own moments, such as variances $\mathrm{Var}(\widehat{\alpha}_i)$ and covariances $\mathrm{Cov}(\widehat{\alpha}_i, \widehat{\alpha}_j)$. In the following we view the estimated EOFs $\widehat{\vec{e}}^{\,i}$ as being 'fixed' (or frozen) rather than random.

Intuitively one would hope that the variance of $\widehat{\alpha}_i$ is equal to that of the real EOF coefficient α_i. Unfortunately, this is not the case (see [13.2.6]). Consider, for example, the first EOF $\vec{e}^{\,1}$ and the corresponding EOF coefficient $\alpha_1 = \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,1} \rangle$. The first EOF minimizes

$$\epsilon_1 = \mathcal{E}\left(\| \vec{\mathbf{X}} - \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,1} \rangle \vec{e}^{\,1} \|^2 \right).$$

Replacing $\vec{e}^{\,1}$ with any other vector, such as $\widehat{\vec{e}}^{\,1}$, increases ϵ_1. Thus

$$\begin{aligned} &\mathrm{Var}(\vec{\mathbf{X}}) - \mathrm{Var}(\alpha_1) \\ &= \mathcal{E}\left(\| \vec{\mathbf{X}} - \langle \vec{\mathbf{X}}, \vec{e}^{\,1} \rangle \vec{e}^{\,1} \|^2 \right) \\ &< \mathcal{E}\left(\| \vec{\mathbf{X}} - \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,1} \rangle \widehat{\vec{e}}^{\,1} \|^2 \right) \\ &= \mathrm{Var}(\vec{\mathbf{X}}) - \mathrm{Var}(\widehat{\alpha}_1), \end{aligned}$$

that is, $\mathrm{Var}(\alpha_1) > \mathrm{Var}(\widehat{\alpha}_1)$. Similar arguments lead to

- $\mathrm{Var}(\widehat{\alpha}_i) < \mathrm{Var}(\alpha_i)$ for the first few EOFs (those corresponding to the largest eigenvalues).

Since the total variance $\mathrm{Var}(\vec{\mathbf{X}}) = \sum_{j=1}^{m} \mathrm{Var}(\vec{\mathbf{x}}_j)$ is estimated with nearly zero bias by $\widehat{\mathrm{Var}}(\vec{\mathbf{X}}) = \sum_{j=1}^{m} \widehat{\lambda}_j$, it follows that

- $\mathrm{Var}(\widehat{\alpha}_i) > \mathrm{Var}(\alpha_i)$ for the last few EOFs.

Examples show that these deviations may be considerable, in particular for small eigenvalues [392].

13.3.3 The Bias in Estimating Eigenvalues. It is natural to ask questions about the reliability of eigenvalues and EOF estimates, such as the extent to which the estimated patterns resemble the true patterns and how close the estimated and true eigenvalues are. These questions do not have completely satisfactory answers, but there are a number of potentially useful facts. One of these facts is the following set of asymptotic formulae that apply to eigenvalue estimates computed from samples that can be represented by n independent and identically distributed normal random vectors (Lawley [245]):

$$\mathcal{E}(\widehat{\lambda}_i) = \lambda_i \left(1 + \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{\lambda_j}{\lambda_i - \lambda_j} \right) + \mathcal{O}(n^{-2}) \tag{13.38}$$

$$\mathrm{Var}(\widehat{\lambda}_i) = \frac{2\lambda_i^2}{n} \left(1 - \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left(\frac{\lambda_j}{\lambda_i - \lambda_j} \right)^2 \right) + \mathcal{O}(n^{-3}). \tag{13.39}$$

As usual, m is the dimension of the random vectors. The symbol $\mathcal{O}(n^{-s})$ represents a term that converges to zero as $n \to \infty$ at least as quickly as n^{-s} does.

By equations (13.38) and (13.39), the eigenvalue estimators are consistent:

$$\lim_{n \to \infty} \mathcal{E}\left((\widehat{\lambda}_i - \lambda_i)^2 \right) = 0. \tag{13.40}$$

However, something unwanted is hidden in equation (13.38), namely that the estimators of the largest and of the smallest eigenvalues are biased. For the largest eigenvalues, almost all of the denominators in equation (13.38) are positive so that the entire sum is positive, that is, $\mathcal{E}(\widehat{\lambda}_i) > \lambda_i$ for large eigenvalues. Similarly, $\mathcal{E}(\widehat{\lambda}_i) < \lambda_i$ for the smallest eigenvalues.

Together with the results from [13.3.2] this finding shows that

- for the largest eigenvalues λ_i,

$$\mathcal{E}(\widehat{\lambda}_i) > \lambda_i = \mathrm{Var}(\alpha_i) > \mathrm{Var}(\widehat{\alpha}_i), \tag{13.41}$$

- for the smallest eigenvalues λ_i,

$$\mathcal{E}(\widehat{\lambda}_i) < \lambda_i = \mathrm{Var}(\alpha_i) < \mathrm{Var}(\widehat{\alpha}_i). \tag{13.42}$$

Relations (13.41) and (13.42) illustrate that we must be cautious when using estimated eigenvalues. First, the estimates are biased: the

large eigenvalues are overestimated and the small ones are underestimated. More important, though, is the inequality $\mathcal{E}(\widehat{\lambda}_i) \neq \mathrm{Var}(\widehat{\alpha}_i)$: The sample eigenvalue $\widehat{\lambda}$ is a biased estimator of the variance of $\widehat{\alpha}_i = \langle \vec{\mathbf{X}}, \widehat{\vec{e}}^{\,i} \rangle$, for any frozen set of estimated EOFs $\widehat{\vec{e}}^{\,i}$. Similarly, $\mathrm{Cov}(\widehat{\alpha}_i, \widehat{\alpha}_j) \neq \widehat{\mathrm{Cov}}(\widehat{\alpha}_i, \widehat{\alpha}_j) = 0$.

13.3.4 'Selection Rules.' Many so-called *selection rules* have been proposed that supposedly separate the physically relevant EOFs from those that are not.[19] One popular procedure of this type is 'Rule N' [321]. The basic supposition is that the full phase space can be partitioned into one subspace that contains only noise and another that contains dynamical variations (or 'signals'). It is assumed that the signal-subspace is spanned by well-defined EOFs while those in the noise-subspace are degenerate. Thus, the idea is to attempt to identify the signal-subspace as the space spanned by the EOFs that are associated with large, well-separated eigenvalues.

The selection rules compare the eigenspectrum[20] estimated from the sample with distributions of sample eigenspectra that are obtained under the assumption that all or the smallest $\widetilde{m}$ true eigenvalues are equal. The number $\widetilde{m}$ is either specified *a priori* or determined recursively. All estimated eigenvalues that are larger than, say, the 95% percentile of the (marginal) distribution of the reference 'noise' spectra, are identified as being 'significant' at the 5% level.

One problem with this approach is that this selection rule is mistakenly understood to be a *statistical test* of the null hypothesis that EOFs $\vec{e}^{\,1}, \ldots, \vec{e}^{\,\widetilde{m}}$, for $\widetilde{m} < m$, span noise against the alternative hypothesis that they span the signal-subspace. The connection between this alternative and the determination of a 'signal-subspace' is vague. Also, the approach sketched above does not consider the reliability of the estimated patterns since the selection rules are focused only on the eigenvalues.

The other problem with the 'selection rule' approach is that there need not be any connection between the shape of the eigenspectrum on the one hand and the presence or absence of 'dynamical structure' on the other. To illustrate, suppose that a process $\vec{\mathbf{X}}_t = \vec{D}_t + \vec{\mathbf{N}}_t$, containing both dynamical and noise components (recall [10.1.1]), has eigenvalues $\lambda_1, \ldots, \lambda_m$ and EOFs $\vec{e}^{\,1}, \ldots, \vec{e}^{\,m}$. Now construct a multivariate white noise $\vec{\mathbf{Z}}_t$ from iid $\mathcal{N}(0, \mathcal{I})$ random vectors. Then the multivariate white noise process $\vec{\mathbf{Y}}_t = \mathcal{P}\Lambda^{1/2}\vec{\mathbf{Z}}_t$, which is completely devoid of 'dynamics,' has the same eigenvalues and eigenvectors as $\vec{\mathbf{X}}_t$. Thus we cannot always diagnose dynamical structure from the zero lag covariance structure of a process.[21]

Our recommendation is to avoid using selection rules. We outline a better approach, based on *North's Rule-of-Thumb*, in the next subsection.

13.3.5 North's Rule-of-Thumb. Using a scale argument, North et al. [296] obtained an approximation for the 'typical' error of the estimated EOFs

$$\Delta\widehat{\vec{e}}^{\,i} \approx \sqrt{\frac{2}{n}} \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{c}{\lambda_j - \lambda_i} \vec{e}^{\,j} \tag{13.43}$$

where c is a constant and n is the number of *independent* samples. There are three things to notice about this equation.

- The first-order error $\Delta\widehat{\vec{e}}^{\,i}$ is of the order of $\sqrt{\frac{1}{n}}$. Thus convergence to zero is slow.
- The first-order error $\Delta\widehat{\vec{e}}^{\,i}$ is orthogonal to the true EOF $\vec{e}^{\,i}$.
- The estimate of the ith EOF $\vec{e}^{\,i}$ is most strongly contaminated by the patterns of those other EOFs $\vec{e}^{\,j}$ that correspond to the eigenvalues λ_j closest to λ_i. The smaller the difference between λ_j and λ_i, the more severe the contamination.

Lawley's formulae (13.38, 13.39) yield a first-order approximation of the 'typical error' in $\widehat{\lambda}_i$:

$$\Delta\lambda_i \approx \sqrt{\frac{2}{n}}\lambda_i. \tag{13.44}$$

Combining this with a simplified version of approximation (13.44), North et al. [296] finally obtain

$$\Delta\widehat{\vec{e}}^{\,i} \approx \frac{c'\Delta\lambda_i}{\lambda_j - \lambda_i} \vec{e}^{\,j} \tag{13.45}$$

where c' is a constant and λ_j is the the closest eigenvalue to λ_i. North's 'Rule-of-Thumb' follows from approximation (13.45): *'If the sampling error*

[19] See, for example, Preisendorfer, Zwiers, and Barnett [321].

[20] An *eigenspectrum* is the distribution of variance (i.e., eigenvalues), with EOF index. The eigenspectrum is an analogue of the power spectrum (see Section 11.2) since both describe the distribution of variance across the coefficients of orthonormal basis functions.

[21] We would need to also analyse at least part of the lagged covariance structure of $\vec{\mathbf{X}}_t$ to reveal the 'dynamics' in this example.

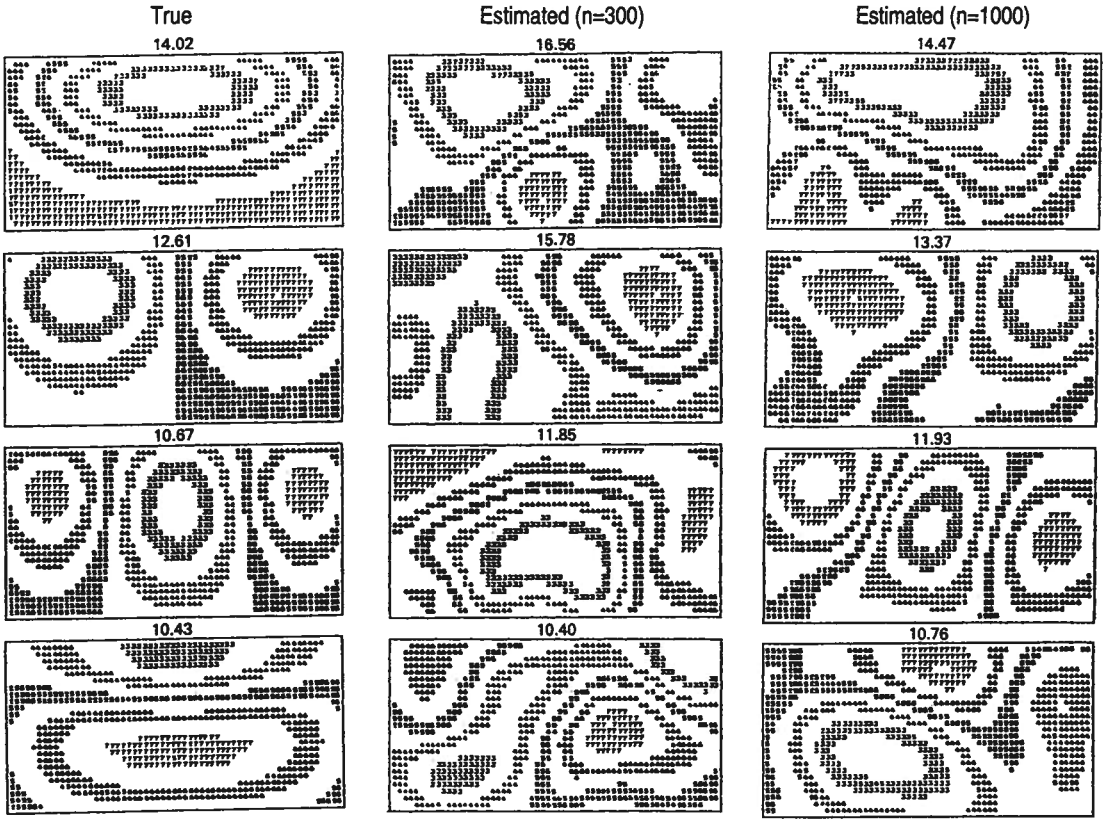

Figure 13.2: *North et al.'s [296] illustration of North's Rule-of-Thumb [13.3.5]. From [296].*
Left: The first four true eigenvalues and EOFs.
Middle: Corresponding estimates obtained from a random sample of size $n = 300$.
Right: As middle column, except $n = 1000$.

of a particular eigenvalue $\Delta\lambda$ *is comparable to or larger than the spacing between* λ *and a neighbouring eigenvalue, then the sampling error* $\Delta\vec{e}$ *of the EOF will be comparable to the size of the neighbouring EOF'.*

13.3.6 North et al.'s Example. North et al. [296] constructed a synthetic example in which the first four eigenvalues are 14.0, 12.6, 10.7 and 10.4 to illustrate North's Rule-of-Thumb [13.3.4]. The first four (true) EOFs are shown in the left hand column of Figure 13.2. According to approximation (13.44) the typical error for the first four estimated eigenvalues is $\Delta\lambda_i \approx \pm 1$ for $n = 300$ and $\Delta\lambda_i \approx \pm 0.6$ for $n = 1000$. Since $\lambda_1 - \lambda_2 = 1.4$, $\lambda_2 - \lambda_3 = 2$ and $\lambda_3 - \lambda_4 = 0.3$, one would expect the first two EOFs to be *mixed*[22] when $n = 300$ and the third and fourth EOF to be mixed for both $n = 300$ and $n = 1000$. That this is a reasonable guess is demonstrated in the middle and right hand columns of Figure 13.2, which displays EOFs estimated from random samples of size $n = 300$ and $n = 1000$, respectively.

[22] That is, we expect the first two EOFs to be a combination of the EOFs that correspond to nearby eigenvalues.

13.4 Examples

13.4.1 Overview. We will present two examples of conventional EOF analysis in this section. This first case, on the globally distributed SST, is most straight forward. The second example involves a data vector that is constructed by combining the same variable at several levels in the vertical.

13.4.2 Monthly Mean Global Sea-surface Temperature. The first two EOFs of monthly mean sea-surface temperature (SST) of the global ocean between 40°S and 60°N are shown in Figure 13.3. They represent 27.1% and 7.9% of the total variance, respectively.

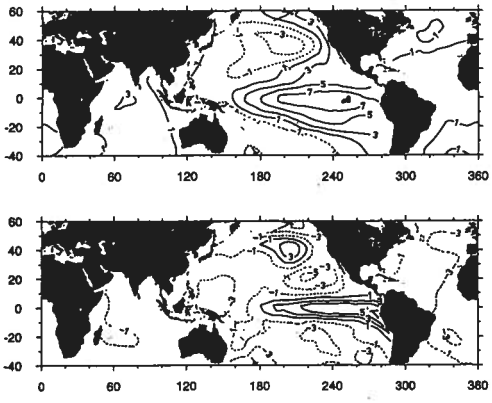

Figure 13.3: *EOFs 1 (top) and 2 (bottom) of monthly mean sea-surface temperature (SST). Units:* 10^{-2}. *Courtesy Xu.*

The first EOF, which is concentrated on the Pacific Ocean, represents ENSO. Its time coefficient, shown as curve 'D' in Figure 13.4, is highly correlated with the two Southern Oscillation indices (Darwin minus Papeete SLP, curve 'E', and SST area average, curve 'F') introduced previously. The large 'centre of action' in the North Pacific represents the oceanic response to anomalous extratropical winds which, in turn, were excited by the anomalous tropical state.

The second EOF of the SST field also involves the tropical Pacific Ocean. The most prominent feature is the narrow tongue of water in the eastern and central equatorial Pacific with temperatures that vary coherently. While the coefficient time series (Figure 13.4, curve 'A') reflects ENSO events (e.g., 1982/83) in part, the connection with the SOI is not as clear as with the first EOF. The coefficient appears to have a downward trend from about 1976 onwards, which would correspond to cooling in the eastern and central tropical Pacific or warming elsewhere. It remains to determine whether the trend is real, part of the global ocean's natural low-frequency variability, or just an artifact of the way in which these data have been collected and analysed.

13.4.3 Monthly Mean of Zonal Wind at Various Levels. The next example is on the monthly mean zonal wind in the troposphere (Xu, personal communication). A joint analysis of the wind field at the 850, 700, 500, 300 and 200 hPa levels was performed. The size of the problem was kept manageable by performing the analysis in two steps. Separate EOF analyses were first performed at each level. In each analysis, the coefficients representing 90% of the variance were retained. A combined vector, composed of EOF coefficients selected for the five levels, is used as input for the eventual EOF analysis of the three-dimensional zonal wind field.

The first two EOFs are shown in Figure 13.5, and their coefficient time series are shown as traces 'B' and 'C' in Figure 13.4. The first EOF, representing 11% of the total monthly variance, is mostly barotropic, not only in the extratropics but also in the tropics. Its coefficient time series exhibits a trend parallel to that found in the coefficient of the second SST EOF. The mean westerly winds in the Southern Hemisphere were analysed as being weaker in the 1970s than in the mid 1980s (negative sign indicates easterly wind anomalies). At the same time the mean low-level easterlies along the equatorial Pacific were weaker in the early 1970s and stronger in the mid 1980s (positive anomalies represent anomalous westerly winds). The results of the EOF analysis of the SST in [13.4.2] are consistent with this representation: The second SST EOF described an equatorial Pacific that was warmer in the early 1970s and cooler in the 1980s, a phenomenon that should be accompanied by strengthening easterly trades during this period.

Because independent analysis techniques are used to derive the SST and zonal wind fields, we can conclude that the trend found in both EOF analyses is not due to data problems. However, it is still not possible to determine whether the trend originates from a natural low-frequency variation or from some other cause.

13.5 Rotation of EOFs

13.5.1 Introduction. This section describes a class of basis vector 'rotation' procedures that is widely used in climate research. The procedures are usually applied to EOFs in the hope that the resulting 'rotated EOFs' can be more easily interpreted than the EOFs themselves. The term 'rotated EOFs' is a mild misnomer that may lead to confusion; 'rotation' transforms the Empirical Orthogonal Function into a *non-orthogonal* linear basis. Also, 'rotation' can be performed on any linear basis, not just EOFs.

We will first explain the general idea and will then describe 'varimax'-rotation in some detail. We use three examples to describe the merits

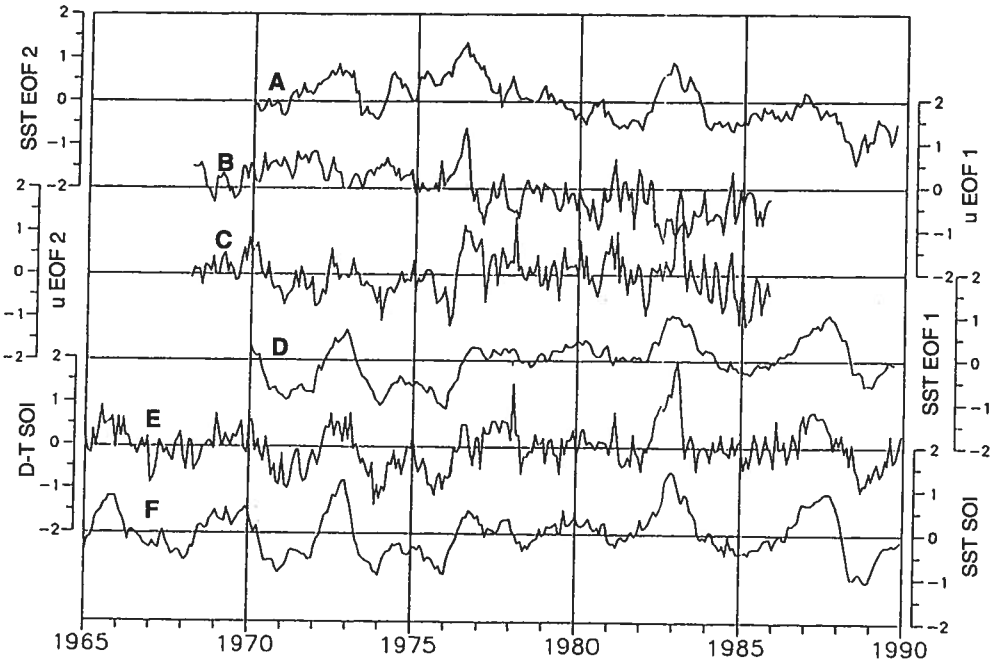

Figure 13.4: *EOF coefficients of monthly mean sea-surface temperature [13.4.2] (curves D and A), monthly mean zonal wind [13.4.3] (curves B and C) and two Southern Oscillation indices [1.2.2] (curves E and F). All data are normalized.*
A: *2nd SST EOF coefficient.* **B**: *1st zonal wind EOF coefficient.* **C**: *2nd zonal wind EOF coefficient.* **D**: *1st SST EOF coefficient.* **E**: *Darwin minus Papeete SLP index of the Southern Oscillation.* **F**: *SST index of the Southern Oscillation.*
Courtesy Xu.

of this procedure. The first of these examples is on the successful and reproducible identification of teleconnection patterns (cf. Section 17.4). The second example deals with a case in which the effect of the rotation is negligible. The third case illustrates pathological behaviour by showing that rotation sometimes splits features into different patterns even though they are part of the *same physical pattern*.

13.5.2 The Concept of 'Rotation.' Having used EOF analysis, or some other technique, to identify a low-dimensional subspace that contains a substantial fraction of the total variance, it is sometimes of interest to look for a linear basis of this subspace with specified properties, such as the following.

- Basis vectors that contain simple geometrical patterns. Simplicity could mean that the patterns are confined regionally, or that the patterns are composed of two regions, one with large positive values and another with large negative values.

- Basis vectors that have time coefficients with specific types of behaviour, such as having nonzero values only during some compact time episodes.

Richman [331] lists five vague criteria for simple structure and there are many proposals of 'simplicity' functionals. The minimization of these functionals is generally non-trivial since the functionals are nonlinear. Numerical algorithms used to obtain approximate solutions can only be applied to bases of moderate size.

The results of a rotation exercise depend on the number and length of the 'input vectors', and on the measure of simplicity. Successful application of the rotation technique requires some experience and the novice may find Richman's [331] review paper on rotation useful. Interesting examples are

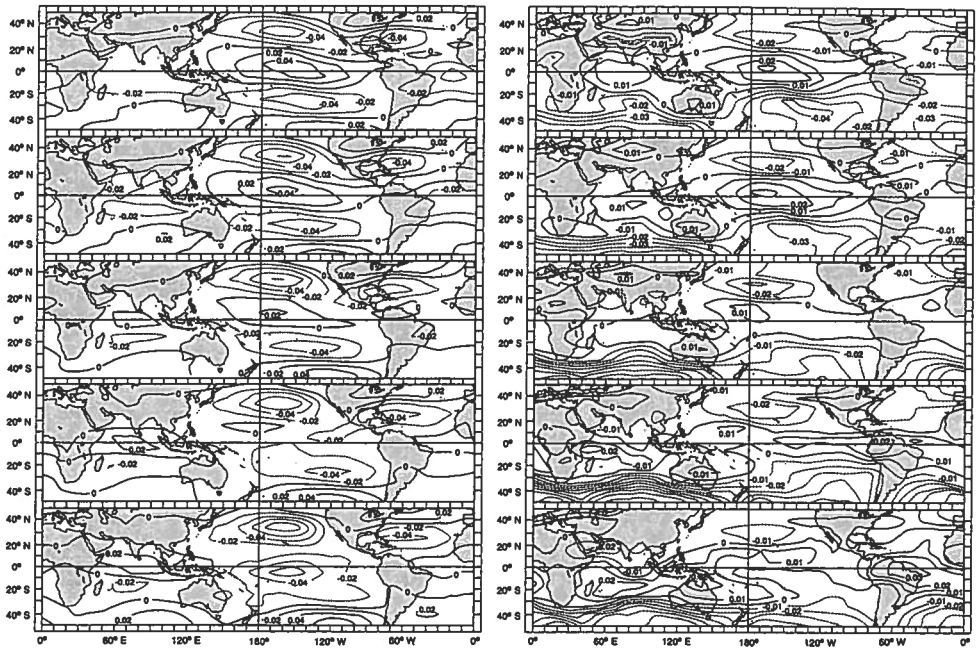

Figure 13.5: *First two EOFs of the tropospheric zonal wind between 45°S and 45°N, at 850, 700, 500, 300 and 200 hPa (from bottom to top). First EOF on the left, second on the right. Courtesy Xu.*

offered by Barnston and Livezey [27] and Chelliah and Arkin [80], among many others.

The opinion in the community is divided on the subject of rotation. Part of the community advocates the use of rotation fervently, arguing that it is a means with which to diagnose physically meaningful, statistically stable patterns from data. Several arguments are raised in favour of the rotated EOFs.

- The technique produces compact patterns that can be used for 'regionalization,' that is, to divide an area in a limited number of homogeneous sub-areas.
- Rotated EOFs are less sensitive to the distribution of observing locations than conventional EOFs.
- Rotated EOFs are often statistically more stable then conventional EOFs (see, e.g., Cheng, Nitsche, and Wallace [82]). That is, the sampling variance of rotated EOFs is often less than that of the input vectors.

Others in the scientific community are less convinced because of the heuristic arguments that motivate the simplicity functionals, and thus the heuristic basis for the interpretation of the result. Jolliffe [198] lists four drawbacks of the routine use of rotation, namely i) the arbitrary choice of the rotation criterion, ii) the sensitivity of the result to the normalization of the EOFs (see [13.5.3]), iii) the need to redo the entire calculation if the number of EOFs is changed (see [13.5.4]), and iv) the loss of information about the dominant sources of variation in the data.

13.5.3 The Mathematics of 'Rotation.' 'Rotation' consists of the transformation of a set of 'input vectors' $\mathcal{P} = (\vec{p}^{\,1}|\cdots|\vec{p}^{\,K})$ into another set of vectors $\mathcal{Q} = (\vec{q}^{\,1}|\cdots|\vec{q}^{\,K})$ by means of an invertible $K \times K$ matrix $\mathcal{R} = (r_{ij})$:

$$\mathcal{Q} = \mathcal{P}\mathcal{R} \tag{13.46}$$

or, for each vector $\vec{q}^{\,i}$:

$$\vec{q}^{\,i} = \sum_{j=1}^{K} r_{ij}\vec{p}^{\,j}. \tag{13.47}$$

The matrix $\mathcal{R}$ is chosen from a class of matrices, such as orthonormal matrices, subject to the

constraint that a functional $V(\mathcal{Q})$ is minimized. An example of such a functional is described in the next subsection.

Under some conditions, operation (13.46) can be viewed as a rotation of the 'input vectors.' Since these are often the first K EOFs, the resulting vectors $\vec{q}^{\,i}$ are called 'rotated EOFs.'

When matrix $\mathcal{R}$ is orthonormal, the operation is said to be an 'orthonormal rotation'; otherwise it is said to be 'oblique.'

Now let $\vec{\mathbf{X}}$ be a random vector that takes values in the space spanned by the input vectors. That is

$$\vec{\mathbf{X}} = \mathcal{P}\vec{\alpha} \tag{13.48}$$

where $\vec{\alpha}$ is a k-dimensional vector of random expansion coefficients. Then, because of operation (13.46)

$$\vec{\mathbf{X}} = (\mathcal{P}\mathcal{R})(\mathcal{R}^{-1}\vec{\alpha}) = \mathcal{Q}\vec{\beta} \tag{13.49}$$

where $\vec{\beta} = \mathcal{R}^{-1}\vec{\alpha}$ is the k-dimensional vector of random expansion coefficients for the rotated patterns.

Let us assume for the following that the matrix $\mathcal{R}$ is orthonormal so that $\vec{\beta} = \mathcal{R}^{\mathrm{T}}\vec{\alpha}$.[23]

- When the input vectors are orthogonal, the scalar products between all possible pairs of rotated vectors are given by the matrix

$$\mathcal{Q}^{\mathrm{T}}\mathcal{Q} = \mathcal{R}^{\mathrm{T}}\mathcal{P}^{\mathrm{T}}\mathcal{P}\mathcal{R} = \mathcal{R}^{\mathrm{T}}\mathcal{D}\mathcal{R}, \tag{13.50}$$

  where $\mathcal{D} = (\vec{p}^{\,1\mathrm{T}}\vec{p}^{\,1}, \ldots, \vec{p}^{\,K\mathrm{T}}\vec{p}^{\,K})$. Thus the rotated vectors are orthogonal only if $\mathcal{D} = \mathcal{I}$, or, in other words, if the input vectors are normalized to unit length.

- Similarly, if the expansion coefficients of the input vectors are pairwise uncorrelated, so that $\Sigma_{\alpha\alpha} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_K^2)$, then the coefficients of the rotated patterns are also pairwise uncorrelated only if coefficients α_j have unit variance. Then

$$\begin{aligned}\Sigma_{\beta\beta} &= \mathrm{Cov}(\mathcal{R}^{\mathrm{T}}\vec{\alpha}, \mathcal{R}^{\mathrm{T}}\vec{\alpha}) \\ &= \mathcal{R}^{\mathrm{T}}\Sigma_{\alpha\alpha}\mathcal{R}.\end{aligned} \tag{13.51}$$

Equations (13.50) and (13.51) imply that rotated patterns derived from normalized EOFs, as defined in [13.1.2,3] so that $\vec{p}^{\,j} = \vec{e}^{\,j}$, are also orthonormal, but their time coefficients are not uncorrelated. If, on the other hand, the EOFs are re-normalized as in equations (13.21) and (13.22) so that $\vec{p}^{\,j} = \vec{e}^{\,j+}$ and $\mathrm{Var}\left(\alpha_j^+\right) = 1$, then the rotated patterns are no longer orthogonal but the coefficients remain pairwise uncorrelated.

Thus two important conclusions may be drawn.

- The result of the rotation exercise depends on the lengths of the input vectors. Differently scaled but directionally identical sets of input vectors lead to sets of rotated patterns that are directionally different from one another. Jolliffe [199] demonstrates that the differences can be large.

  The rotated vectors are a function of the input vectors rather than the space spanned by the input vectors.

- After rotating EOF patterns, the new patterns and coefficients are not orthogonal and uncorrelated at the same time. When the coefficients are uncorrelated, the patterns are not orthogonal, and vice versa. Thus, the percentage of variance represented by the individual patterns is no longer additive.

13.5.4 The 'Varimax' Method. 'Varimax' is a widely used orthonormal rotation that minimizes the 'simplicity' functional

$$V(\vec{q}^{\,1}, \ldots, \vec{q}^{\,K}) = \sum_{i=1}^{K} f_V(\vec{q}^{\,i}) \tag{13.52}$$

where $\vec{q}^{\,i}$ is given by equation (13.47) and f_V is defined by

$$f_V(\vec{q}\,) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{q_i}{s_i}\right)^4 - \frac{1}{m^2}\left(\sum_{i=1}^{m}\left(\frac{q_i}{s_i}\right)^2\right). \tag{13.53}$$

The constants s_i are chosen by the user. The *raw varimax* rotation is obtained when $s_i = 1$, $i = 1, \ldots, K$, and the *normal varimax* rotation is obtained by setting $s_i = \sum_{j=1}^{K}(p_i^j)^2$. Another option is to define s_i as the standard deviation of the ith component of

$$\vec{\mathbf{X}}^{(K)} = \sum_{j=1}^{K}\alpha_j\vec{p}^{\,j},$$

which is the projection of the original full random vector $\vec{\mathbf{X}}$ onto the subspace spanned by the K vectors $\{\vec{p}^{\,1} \ldots \vec{p}^{\,K}\}$.

Note that $f_V(\vec{q}\,)$ (13.53) can be viewed as the spatial variance of the normalized squares $(q_i/s_i)^2$. That is, $f_V(\vec{q}\,)$ measures the 'weighted

[23] All matrices and vectors in this section are assumed to be real valued. Thus orthonormal matrices satisfy $\mathcal{R}\mathcal{R}^{\mathrm{T}} = \mathcal{R}^{\mathrm{T}}\mathcal{R} = \mathcal{I}$.

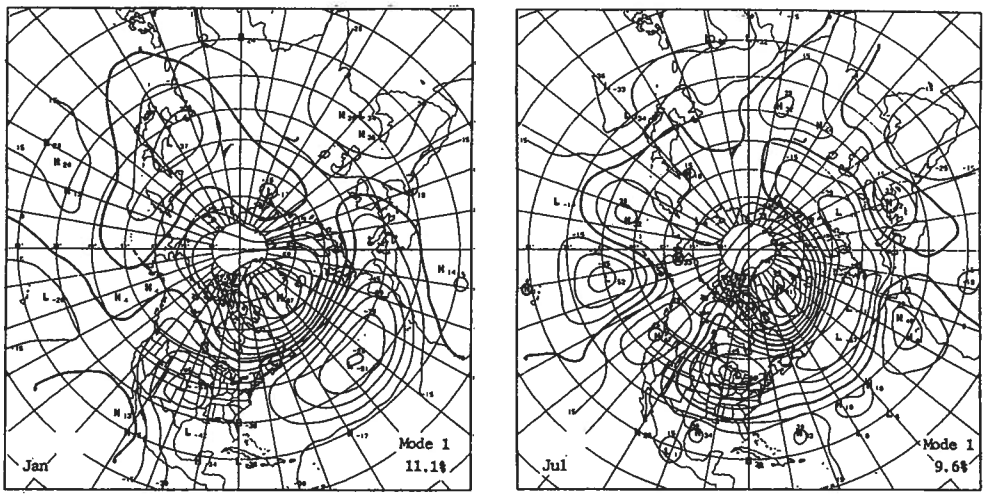

Figure 13.6: *January (left) and July (right) versions of the North Atlantic Oscillation pattern derived by Barnston and Livezey [27] by applying varimax rotation to the first 10 normalized EOFs of January and July mean 700 hPa height, respectively. Courtesy R. Livezey.*

square amplitude' variance of $\vec{q}$. Therefore, minimizing function (13.52) is equivalent to finding a matrix $\mathcal{R}$ such that the sum of the total weighted square amplitude variance of the K patterns $(\vec{q}^{\,1}|\cdots|\vec{q}^{\,k}) = \mathcal{P}\mathcal{R}$ is minimized. See Richman [331] for further details.

13.5.5 Example: Low-frequency Atmospheric Circulation Patterns. Barnston and Livezey [27] argued extensively that rotated EOF analysis is a more effective tool for the analysis of atmospheric circulation patterns than the 'teleconnection' analysis (Wallace and Gutzler [409]; see also Section 17.4). They used a varimax rotation of re-normalized EOFs (13.21, 13.22) to isolate the dominant circulation patterns in the Northern Hemisphere (NH) on the monthly time scale. The EOFs used in the study were computed separately for each month of the year from correlation matrices derived from a 35-year data set of monthly mean 700 hPa heights analysed on a 358-point grid. The data set itself was carefully screened to remove known analysis biases. Rotation was performed on the first 10 EOFs in each month. They represent about 80% of the total variance in winter and 70% in summer.

The result of the exercise is an extensive collection of NH circulation patterns. Barnston and Livezey identified 13 patterns: nine cold season patterns, two warm season patterns, and two transition season patterns. Only one pattern, the North Atlantic Oscillation (NAO, Figure 13.6) is evident in every month of the year. Barnston and Livezey estimate that it represents between 15.4% (March) and 7.4% (October) of the total variance. The NAO is the dominant circulation pattern in the solstitial seasons (DJFM and MJJAS). The NAO is characterized by a 'high' (this adjective is arbitrary since the sign of the pattern is arbitrary) that is centred, roughly, over Greenland and a low pressure band to the south. Figure 13.6 displays 'typical' configurations in winter and summer. The Greenland centre is located at about 70° N and 40–60° W in winter, and has a zero line at about 50° N. This centre retreats northward in summer and a second zero line appears at about 30–35° N.

Another pattern extracted by Barnston and Livezey that has been studied by many others is the Pacific/North American (PNA) pattern (Figure 13.7). The PNA is characterized by two centres of the same sign over the Aleutian Islands and the southeastern United States that flank a centre of opposite sign located over western North America. The PNA is evident in winter (December to April) and again in September and October. It is strongest in February when Barnston and Livezey estimate that it represents 13.2% of the total variance.

Even though the rotated EOFs appear to be less prone to 'mixing' than ordinary EOFs, a great deal of sampling variability still clouds the patterns that are produced, and a considerable

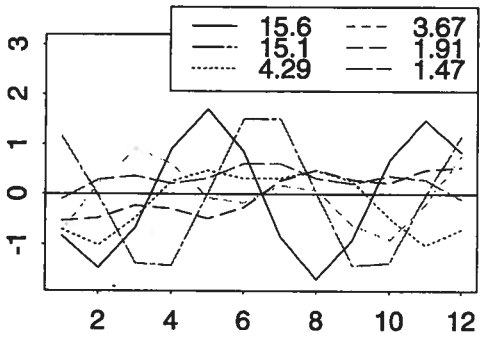

Figure 13.13: *First six time EOFs of an AR(2) process with* $\vec{a} = (0.9, -0.8)$ *obtained using window length* $m = 12$. *The patterns are normalized with the square root of the eigenvalue. The eigenvalues are given at the bottom.*

13.6.8 Multichannel Singular Spectrum Analysis. MSSA (see Vautard [380]) differs from SSA only in the dimension of the basic time series, which is now m'-dimensional rather than one-dimensional. The derived random vector $\vec{\mathbf{Y}}_t$ is therefore mm'-dimensional. Thus MSSA is *Extended EOF analysis* [13.1.8] in which m consecutively observed fields are concatenated together. The number of fields m is usually small compared with the field dimension m' in EEOF analysis. The opposite, $m > m'$, is often true in MSSA.

13.6.9 Estimation. We conclude with a brief comment on the estimation of eigenvalues and time EOFs in SSA. The same applies, by extension, to the eigenvalues and space-time EOFs in MSSA.

SSA is applied to a finite sample of observations $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ with $n \gg m$ by first forming $\vec{\mathbf{Y}}$-vectors,

$$\begin{aligned} \vec{\mathbf{y}}_1 &= (\mathbf{x}_1, \ldots, \mathbf{x}_m)^{\mathrm{T}} \\ \vec{\mathbf{y}}_2 &= (\mathbf{x}_2, \ldots, \mathbf{x}_{m+1})^{\mathrm{T}} \\ &\vdots \\ \vec{\mathbf{y}}_{n-m+1} &= (\mathbf{x}_{n-m+1}, \ldots, \mathbf{x}_n)^{\mathrm{T}}. \end{aligned}$$

Conventional EOF analysis is applied to the resulting sample of $n - m + 1$ $\vec{\mathbf{Y}}$-vectors. The estimated eigenvalues and EOFs can be computed from either the estimated covariance matrix of $\vec{\mathbf{Y}}$ (see [13.2.4]) or by means of SVD (see [13.2.8]).

Note that neither *North's Rule-of-Thumb* [13.3.5] nor Lawley's formulae (13.38, 13.39) can be used to assess the reliability of the estimate directly because consecutive realizations of $\vec{\mathbf{Y}}_t$ are auto-correlated (see (13.33)). The effects of temporal dependence must be accounted for (see Section 17.1) when using these tools.

Allen and co-workers [8, 9, 10] discuss the problem of discriminating between noisy components and truly oscillatory modes in detail.

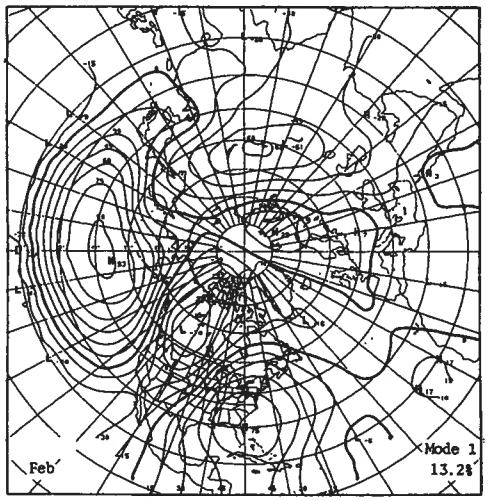

Figure 13.7: *As Figure 13.6, except the February Pacific/North American pattern is displayed. Courtesy R. Livezey.*

amount of skill and subjective judgement are needed to classify and name the patterns. This is amply illustrated by Barnston and Livezey [27], who discuss the types of latitude they permitted themselves in developing their classification. Their illustration of six renditions of the NAO obtained for different times of the year (we show two of these in Figure 13.6; see Barnston and Livezey [27, Figure 2]) demonstrates the kind of variability that the analyst must be able to penetrate when classifying estimated patterns.

13.5.6 Example: Atlantic Sea-Level Air Pressure. In this subsection we consider the EOFs and varimax-rotated EOFs of North Atlantic monthly mean SLP in DJF.[24]

The first three EOFs of SLP (Figure 13.8, left hand column) represent 41%, 26% and 9% of the total variance, respectively. The first EOF has almost uniform sign and exhibits one large feature. The second and third EOFs have dipole structures that reflect the constraint that the higher-order EOFs must be orthogonal to the first EOF.

The EOFs of North Atlantic SLP have simple structure, even without rotation. It is therefore not surprising that the application of the varimax rotation technique to these EOFs (Figure 13.8, middle and right hand columns) results in little change.

- Figure 13.8 (middle column) displays the result of the rotation using $K = 5$ normalized EOFs as input vectors. The rotated EOFs represent 38%, 24% and 10% of the total variance. Similar results are obtained when $K = 5$ non-normalized EOFs are used (not shown).

- The result of the rotation using the first $K = 10$ *non-normalized* EOFs is shown in the right hand column of Figure 13.8. The patterns represent 26%, 15%, and 13% of the total variance, respectively. These patterns deviate somewhat from those in the left hand and middle columns of the diagram. They are noisier than the other sets of patterns, including the rotated patterns derived from $K = 10$ normalized EOFs (not shown).

  Intuitively this is what we expect since the non-normalized patterns enter the minimization functional with equal weights. Thus the poorly estimated EOFs are as influential as the well-estimated EOFs in the determination of matrix $\mathcal{R}$. In contrast, normalization gives the well-estimated EOFs relatively more influence on the form of $\mathcal{R}$.

The first rotated pattern represents less variance than the first EOF, simply because the first EOF was constructed to maximize the variance. Higher-order rotated EOFs typically represent more variance than the respective EOFs (see, e.g., Table 1 of Barnston and Livezey [27]).

In this example, little is gained by processing the original EOF patterns with the varimax machinery. The rotated EOFs become noisy when too many non-normalized EOFs were used as input.

13.5.7 Example: North Atlantic Sea-surface Temperature. The first three EOFs of the monthly mean SST in DJF represent 26%, 17% and 10% of the total variance, respectively (Figure 13.9, left hand column). These EOFs do not have simple structure. The first contains three well-separated centres of location located in the West Atlantic off the North American coast, south of Greenland, and in the upwelling region off the west coast of Africa.

Varimax rotation leads to a substantially different distribution of variance between patterns (Figure 13.9, middle and right hand columns).

[24] The analysis presented here and in [13.5.7] were provided by V. Kharin (personal communication). Note that all eigenvalues, EOFs, and rotated EOFs presented here are estimates.

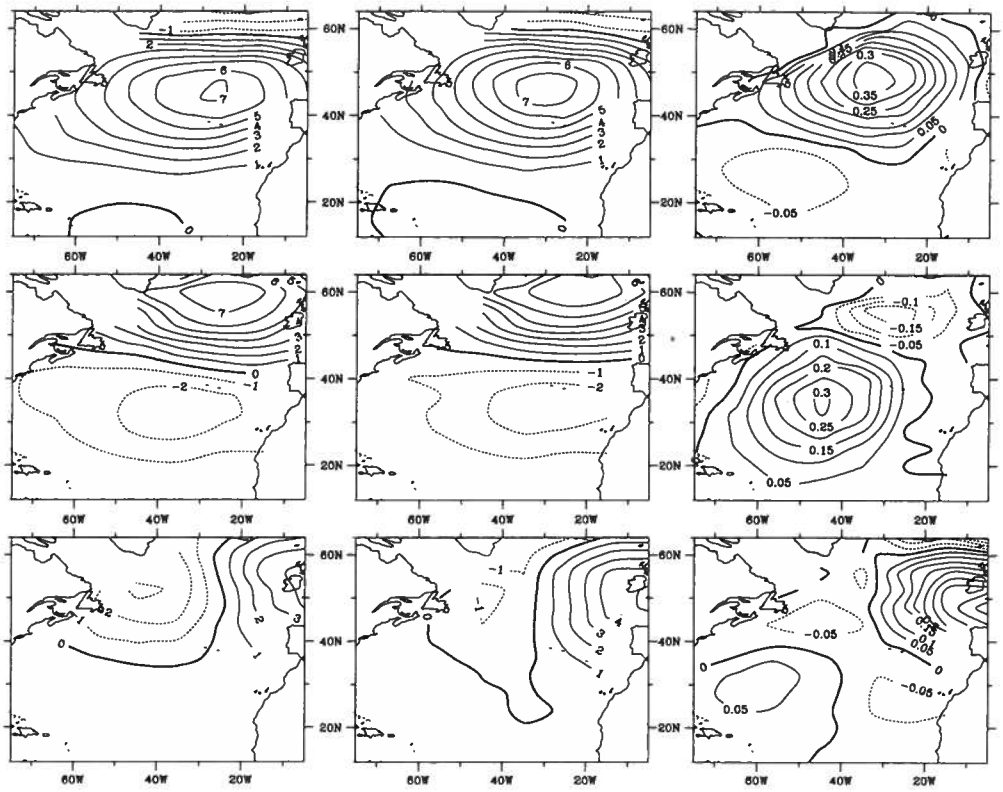

Figure 13.8: *First three rotated and unrotated EOFs of North Atlantic SLP in winter. From top to bottom* $j = 1$, $j = 2$, $j = 3$. *Courtesy V. Kharin.*
Left column: Normalized EOFs $\vec{e}^{\,j+}$.
Middle column: Rotated EOFs derived from $K = 5$ *normalized EOFs.*
Right column: Rotated EOFs derived from $K = 10$ *non-normalized EOFs.*

- When the input is $K = 5$ normalized EOFs (Figure 13.9, middle column) the three centres of action in the first EOF are separated and distributed to the first three rotated patterns (which represent 21%, 16% and 15% of the variance, respectively).

- When the input is $K = 5$ non-normalized EOFs (i.e., all EOFs have unit length; Figure 13.9, right hand column) the three rotated EOFs represent about the same percentage of variance, namely 15%, 15% and 13%, respectively.[25] Note that the sequence of patterns is changed from that obtained with the normalized EOFs (which have unequal lengths). When more input vectors are used, the rotated patterns become noisier and represent less variance (not shown).

We will revisit the analysis of North Atlantic SST and SLP in [14.3.1]. There we will see that the first two conventional EOFs of the North Atlantic SST reflect *two* forcing mechanisms, two characteristic variations in the large-scale atmospheric state that are encoded in the first two SLP EOFs shown in Figure 13.8. Thus, in this case, the rotation makes interpretation more difficult by masking the underlying physics (see [14.3.2]).

[25] Note that the concept of degeneracy is irrelevant for rotated EOFs, since degeneracy is immaterial for the minimization of the functional V.

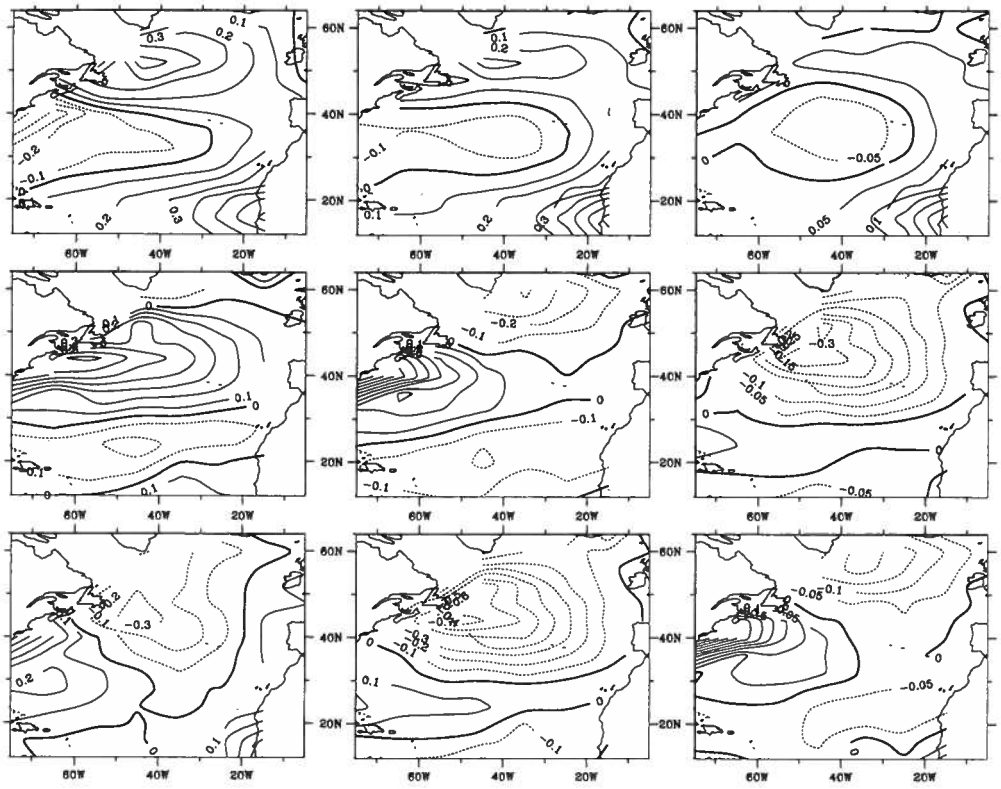

Figure 13.9: *First three unrotated (left hand column) and rotated (middle and right hand columns) EOFs of North Atlantic monthly mean SST in DJF. From top to bottom:* $j = 1$, $j = 2$, $j = 3$. *Courtesy V. Kharin.*
Left column: Normalized EOFs $\vec{e}^{\,j+}$.
Middle column: Rotated EOFs derived from $K = 5$ *normalized EOFs.*
Right column: Rotated EOFs derived from $K = 5$ *non-normalized EOFs.*

13.5.8 Rotation: a Postscript. EOF rotation is often useful, but it is not meant to be a default operation after every EOF analysis. Instead its use should be guided by the problem under consideration.

Jolliffe [198] points out that rotation should be used routinely for subsets of EOFs that have equal, or near-equal, eigenvalues. The corresponding EOFs are not well defined because of their degeneracy (cf. [13.1.8]), and thus the patterns contained by the degenerate EOFs may be arbitrarily rotated within the space that they span. The sensitivity of the rotation to the normalization of the EOFs becomes less relevant since all eigenvalues are similar.

13.6 Singular Systems Analysis and Multichannel SSA

13.6.1 General. The *Singular Systems Analysis* (SSA; see Vautard, Yiou, and Ghil [381] or Vautard [380]) and the *Multichannel Singular Spectrum Analysis* (MSSA, see Plaut and Vautard [317]) are time series analysis techniques used to identify recurrent patterns in univariate time series (SSA) and multivariate time series (MSSA). Mathematically, SSA and MSSA are variants of conventional EOF analysis, but the application of the mathematics is markedly different. Vautard [380] reviews recent applications of SSA and MSSA. Allen and colleagues [8, 9,

10] have investigated various aspects of these methods.

13.6.2 Singular Systems Analysis. Univariate time series $\mathbf{X}_t$ are considered in SSA. An m-dimensional vector time series $\vec{\mathbf{Y}}_t$ is derived from $\mathbf{X}_t$ by setting:

$$\vec{\mathbf{Y}}_t = (\mathbf{X}_t, \mathbf{X}_{t+1}, \ldots, \mathbf{X}_{t+m-1})^{\mathrm{T}}. \tag{13.54}$$

A *Singular Systems Analysis* is an EOF analysis of $\vec{\mathbf{Y}}_t$.

The vector space occupied by $\vec{\mathbf{Y}}_t$ is called the *delay-coordinate space*.

The (zero lag) covariance matrix of $\vec{\mathbf{Y}}$, $\Sigma_{YY} = \mathrm{Cov}(\vec{\mathbf{Y}}_t, \vec{\mathbf{Y}}_t)$, is a *Töplitz matrix*.[26] Element (j, k) of Σ_{YY}, say σ_{jk}, is the covariance between the jth element of $\vec{\mathbf{Y}}_t$ ($\mathbf{X}_{t+j-1}$) and its kth element $\mathbf{X}_{t+k-1}$. Thus

$$\sigma_{jk} = \gamma_{xx}(|j-k|),$$

where $\gamma_{xx}(\cdot)$ is the auto-correlation function of $\mathbf{X}_t$. All off-diagonal elements of Σ_{YY} are identified by $|i-j| = \tau$ and have the same value $\gamma_{xx}(\tau)$. Thus, matrix Σ_{YY} is band-structured and contains all auto-covariances of $\mathbf{X}_t$ up to lag $m-1$. The covariance and correlation matrices of $\vec{\mathbf{Y}}_t$ differ by only a constant factor $(1/\sigma_X^2)$. They therefore have the same eigenvectors. The eigenvalues of the two matrices differ by the same constant factor.

The eigenvectors $\vec{e}^{\,i}$ of Σ_{YY}, sometimes called *time EOFs*, are interpreted as a sequence in time. Each eigenvector $\vec{e}^{\,i}$ is a normalized sequence of m time-ordered numbers,

$$\vec{e}^{\,i} = \left(e_0^i, \ldots, e_{m-1}^i\right)^{\mathrm{T}}, \tag{13.55}$$

that may be understood as a 'typical' sequence of events. The orthogonality of the eigenvectors in the delay-coordinate space, $\vec{e}^{\,j\mathrm{T}}\vec{e}^{\,k} = \delta_{jk}$, is equivalent to the *temporal orthogonality* of any two typical sequences $(e_0^j, \ldots, e_{m-1}^j)$ and $(e_0^k, \ldots, e_{m-1}^k)$:

$$\sum_{i=0}^{m-1} e_i^j e_i^k = \delta_{jk}. \tag{13.56}$$

The EOF coefficients

$$\alpha_k(t) = \langle \vec{\mathbf{y}}_t, \vec{e}^{\,k} \rangle = \sum_{i=0}^{m-1} \mathbf{X}_{t+i} e_i^k \tag{13.57}$$

are empirically determined averages (recall Section 10.5) of length m. That is, $\alpha_k(t)$ is a *filtered*[27] version of the original time series $\mathbf{X}_t$, with filter weights that are given by the kth eigenvector. When $\mathbf{X}_t$ is dominated by high-frequency variations, the dominant eigenvectors will be high-pass filters, and when most of the variance of $\mathbf{X}_t$ is concentrated at low frequencies the dominant eigenvectors will act as low-pass filters. The eigenvectors will generally *not* form symmetric filters. Thus we need to be aware that operation (13.57) causes a frequency-dependent phase shift.

As with ordinary EOF analysis, SSA distributes the total variance of $\vec{\mathbf{Y}}_t$ to the m eigenvalues λ_i. The total variance of $\vec{\mathbf{Y}}_t$ is equal to m times the variance of $\mathbf{X}_t$. Thus

$$\sum_{i=1}^{m} \lambda_i = m\,\mathrm{Var}(\mathbf{X}_t). \tag{13.58}$$

The vector-matrix version of (13.57) is

$$\vec{\alpha}(t) = \mathcal{P}\vec{\mathbf{Y}}_t$$

where $\vec{\alpha}(t)$ and $\mathcal{P}$ are defined in the usual way. Thus the auto-correlation function of the multivariate coefficient process $\vec{\alpha}(t)$ is related to the auto-correlation function of $\mathbf{X}_t$ by

$$\Sigma_{\alpha\alpha}(\tau) = \mathcal{P}\Sigma_{YY}(\tau)\mathcal{P}^{\mathrm{T}}$$

where $\Sigma_{YY}(\tau)$ is the matrix whose (i, j)th entry is given by

$$[\Sigma_{YY}(\tau)]_{i,j} = \gamma_{xx}(\tau + j - i). \tag{13.59}$$

Note that

$$\Sigma_{\alpha\alpha}(0) = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m).$$

13.6.3 Reconstruction in the Time Domain. Also, as with ordinary EOFs,

$$\vec{\mathbf{Y}}_t = \sum_{i=1}^{m} \alpha_i(t)\vec{e}^{\,i}. \tag{13.60}$$

Thus, using equation (13.60) to expand $\vec{\mathbf{Y}}_t, \vec{\mathbf{Y}}_{t-1}, \ldots, \vec{\mathbf{Y}}_{t-m+1}$, we find that $\mathbf{X}_t$ has m equivalent time expansions in the m 'SSA-signals':

$$\begin{aligned} \mathbf{X}_t &= \sum_{i=1}^{m} \alpha_i(t) e_1^i \\ &= \sum_{i=1}^{m} \alpha_i(t-1) e_2^i \\ &\vdots \\ &= \sum_{i=1}^{m} \alpha_i(t-m+1) e_m^i. \end{aligned} \tag{13.61}$$

[26] The elements on each diagonal of a Töplitz matrix are equal. That is, if $\mathcal{A}$ is an $m \times m$ matrix and if there are constants $c_{(-(n-1))}, \ldots, c_{(n-1)}$ such that $\mathcal{A}_{i,j} = c_{j-i}$, then $\mathcal{A}$ is Töplitz. Graybill [148] describes some of their properties (see Section 8.15).

[27] See Section 17.5.

Each of these expansions distributes the variance of the SSA-signals differently. In fact, using the orthogonality of the EOF coefficients, it is easily shown that

$$\mathrm{Var}(\mathbf{X}_t) = \sum_{i=1}^{m} \lambda_i \left(e_k^i\right)^2 \tag{13.62}$$

for *all* k.[28] If we consider the normalized representation (13.22) we find that the SSA patterns add to the same numbers:

$$\sum_i (e_k^{i+})^2 = \text{constant}, \tag{13.63}$$

for all lags k.

13.6.4 Paired Eigenvectors and Oscillatory Components. We now consider, briefly, time series that contain an oscillatory component. For simplicity, we suppose that $\mathbf{X}_t$ is pure cosine so that

$$\begin{aligned}\vec{\mathbf{Y}}_t &= (\mathbf{X}_t, \ldots, \mathbf{X}_{t+m-1})^{\mathrm{T}} \\ &= \left(\cos\left(2\pi\tfrac{t}{m}\right), \ldots, \cos\left(2\pi\tfrac{t+m-1}{m}\right)\right)^{\mathrm{T}}.\end{aligned} \tag{13.64}$$

By equation (13.60), the time EOFs must be able to represent this structure. Suppose one of the time EOFs contains the cosine pattern, that is,

$$\vec{e}^{\,i} = \left(1, \cos\left(\tfrac{2\pi}{m}\right), \ldots, \cos\left(\tfrac{2\pi(m-1)}{m}\right)\right)^{\mathrm{T}}.$$

Then $\vec{\mathbf{Y}}_0 = \vec{e}^{\,i}$. However, one time step later, we have

$$\begin{aligned}\vec{\mathbf{Y}}_1 &= \left(\cos\left(\tfrac{2\pi}{m}\right), \ldots, \cos\left(\tfrac{2\pi(m-1)}{m}\right), 1\right)^{\mathrm{T}} \\ &= \cos\left(\tfrac{2\pi}{m}\right)\left(1, \cos\left(\tfrac{2\pi}{m}\right), \ldots, \cos\left(\tfrac{2\pi(m-1)}{m}\right)\right)^{\mathrm{T}} \\ &\quad - \sin\left(\tfrac{2\pi}{m}\right)\left(0, \sin\left(\tfrac{2\pi}{m}\right), \ldots, \sin\left(\tfrac{2\pi(m-1)}{m}\right)\right)^{\mathrm{T}} \\ &= \cos\left(\tfrac{2\pi}{m}\right)\vec{e}^{\,i} - \sin\left(\tfrac{2\pi}{m}\right)\vec{e}^{\,j},\end{aligned}$$

where $\vec{e}^{\,j} = \left(0, \sin\left(\tfrac{2\pi}{m}\right), \ldots, \sin\left(\tfrac{2\pi(m-1)}{m}\right)\right)^{\mathrm{T}}$ is another eigenvector of $\vec{\mathbf{Y}}_t$. At time t,

$$\begin{aligned}\vec{\mathbf{Y}}_t &= \cos\left(\tfrac{2\pi t}{m}\right)\vec{e}^{\,i} - \sin\left(\tfrac{2\pi t}{m}\right)\vec{e}^{\,j} \\ &= \alpha_i(t)\vec{e}^{\,i} + \alpha_j(t)\vec{e}^{\,j}\end{aligned}$$

where $\alpha_i(t) = \cos\left(\tfrac{2\pi t}{m}\right)$ and $\alpha_j(t) = \sin\left(\tfrac{2\pi t}{m}\right)$. Note that both coefficients have the same 'variance' (i.e., $\lambda_i = \lambda_j$), and that the coefficients are 90° out-of-phase.

While the example is artificial, the properties of the eigenvectors and coefficients above characterize what happens when $\mathbf{X}_t$ contains an oscillatory signal. That is, we expect to find a pair of degenerate EOFs with coefficients that vary coherently and are 90° out-of-phase with each other.[29] The pair of patterns and their coefficients may be written as one complex pattern and one complex coefficient.

[28] Note that (13.62) is just a special case of (13.7).

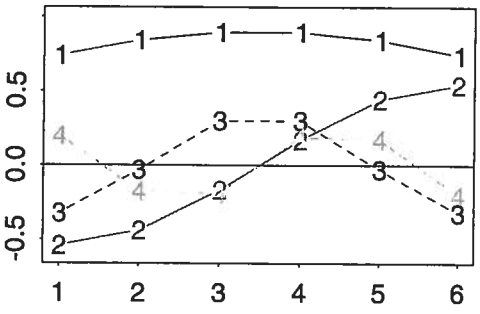

Figure 13.10: *The first four time EOFs of an AR(1) process with* $a = 0.8$ *obtained using window length* $m = 6$. *The patterns are normalized with the square root of the eigenvalue.*

13.6.5 SSA of White Noise. A white noise process $\{\mathbf{X}_t\}$ (see [10.2.3]) consists of a sequence of independent, identically distributed random variables. It has auto-covariance function $\gamma_{xx}(\tau)$ such that $\gamma_{xx}(0) = \mathrm{Var}(\mathbf{X}_t)$ and $\gamma_{xx}(\tau) = 0$ for nonzero τ. Thus

$$\mathbf{\Sigma}_{YY} = \mathrm{Var}(\mathbf{X}_t)\mathcal{I},$$

where $\vec{\mathbf{Y}}_t$ is the delay-coordinate space version of $\mathbf{X}_t$ (13.58) and $\mathcal{I}$ is the $m \times m$ identity matrix. Hence $\vec{\mathbf{Y}}_t$ has m eigenvalues $\lambda_i = \mathrm{Var}(\mathbf{X}_t)$ and m degenerate eigenvectors. One possible set of eigenvectors are the unit vectors, $\vec{e}^{\,i} = (0, \ldots, 1, \ldots, 0)$ with the 1 in the ith column.

13.6.6 SSA of Red Noise. Red noise processes[30] have exponentially decaying auto-covariance functions $\gamma_{xx}(\tau) = \sigma_X^2 a^{|\tau|}$, where $\sigma_X^2 = \mathrm{Var}(\mathbf{X}_t)$. Thus

$$\mathbf{\Sigma}_{YY} = \sigma_X^2 \begin{pmatrix} 1 & a & \ldots & a^{m-1} \\ a & 1 & \ldots & a^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ a^{m-1} & a^{m-2} & \ldots & 1 \end{pmatrix}.$$

[29] Compare with the discussion of complex POP coefficients in Chapter 15.

[30] AR(1), or 'red noise,' processes were introduced in [10.3.2]. They can be represented by a stochastic difference equation $\mathbf{X}_t = a\mathbf{X}_{t-1} + \mathbf{Z}_t$, where $\mathbf{Z}_t$ is white noise. The auto-covariance function was derived in [11.1.6]. We represent the lag-1 correlation coefficient by 'a' instead of 'α' to avoid confusion with our notation for the EOF coefficients.

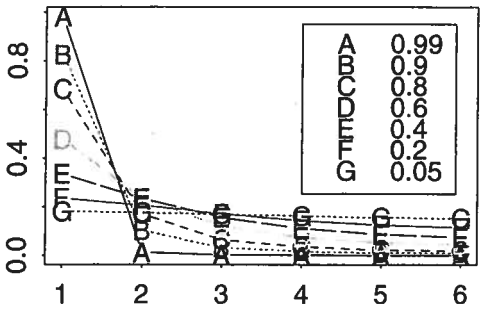

Figure 13.11: *Eigenspectra obtained with window length* $m = 6$ *for AR(1) processes with* $a = 0.99, 0.90, 0.80, 0.60, 0.40, 0.20,$ *and* 0.05. *The spectra are normalized by the variance of the process.*

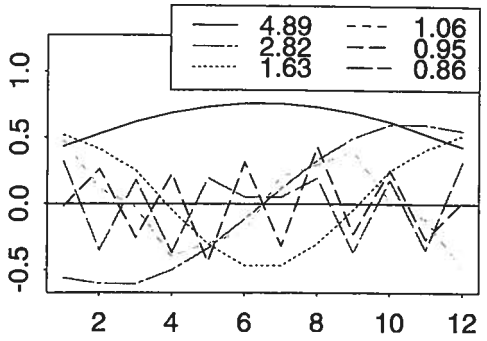

Figure 13.12: *First six time EOFs of an AR(2) process with* $\vec{\alpha} = (0.3, 0.3)$ *obtained using window length* $m = 12$. *The patterns are normalized with the square root of the eigenvalue. The eigenvalues are given at the bottom.*

When $m = 2$, Σ_{YY} has eigenvalues $(1 - a)$ and $(1 + a)$ and corresponding time EOFs $(1/\sqrt{2}, -1/\sqrt{2})^{\mathrm{T}}$ and $(1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$. The order of the eigenvalues and EOFs depends upon the sign of a. Given a specific *window length* m, the same EOFs are obtained for all AR(1) processes $\mathbf{X}_t$.

The first four AR(1) time EOFs for window length $m = 6$ are shown in Figure 13.10. The patterns are multiplied by the square root of the eigenvalue, as in equation (13.22). The kth pattern crosses the zero line $k - 1$ times. Thus, the time EOFs are ordered by time scale, with most variance contributed by the variability with longest time scales.

One characteristic of the time EOFs is that no two patterns have the same number of zeros. Thus oscillatory behaviour, such as that described in [13.6.4], is not possible. This is consistent with the discussion in [11.1.2], when we also found no indication of oscillatory behaviour in AR(1) processes.

Figure 13.11 shows the eigenspectra of several AR(1) processes for the same window length m. The larger the 'memory' a, the steeper the spectrum. In the extreme case with $a = 0.99$, almost all variance is contributed by the 'almost constant' first time EOF. At the other end of the memory scale ($a = 0.05$) all time EOFs contribute about the same amount of variance.

13.6.7 SSA of an AR(2) process. An AR(2) process has the form

$$\mathbf{X}_t = a_1\mathbf{X}_{t-1} + a_2\mathbf{X}_{t-2} + \mathbf{Z}_t$$

[10.3.3]. The auto-covariance function is either the sum of the auto-covariance functions of two red noise processes (11.4), or it is a damped oscillatory function (11.9) (for details, see [11.1.9]).

The process with coefficients $a_1 = a_2 = 0.3$ was found to belong to the former 'non-oscillatory' group. The first six time EOFs obtained using window length $m = 12$ are shown in Figure 13.12. Similar to the AR(1) process, all eigenvectors have different patterns, with the kth eigenvector having $(k - 1)$ zeros. All eigenvalues are well separated. Consistent with the discussion in [11.1.7] and [11.2.6], an oscillatory mode is not identified. Note the similarity between the patterns in Figure 13.12 and the AR(1) patterns shown in Figure 13.10. (The patterns in Figure 13.12 are not sensitive to the choice of m.)

The other AR(2) process considered previously has $\alpha_1 = 0.9, \alpha_2 = -0.8$. This process has oscillatory behaviour with a 'period' of 6 time steps (see [10.3.3], [11.1.7] and [11.2.6]). The time EOFs of this process, obtained using window length $m = 12$, are shown in Figure 13.13. The first two time EOFs are sinusoidal, with a period of 6 time steps, and phase-shifted by 1 to 2 time steps (a quarter of a period, 1.5 time steps, can not be represented in time steps of 1). The two time EOFs share similar eigenvalues (4.2 and 4.1) and obviously represent an oscillatory mode as described in [13.6.4]. The higher index time EOFs are reminiscent of the time EOFs obtained for AR(1) processes.

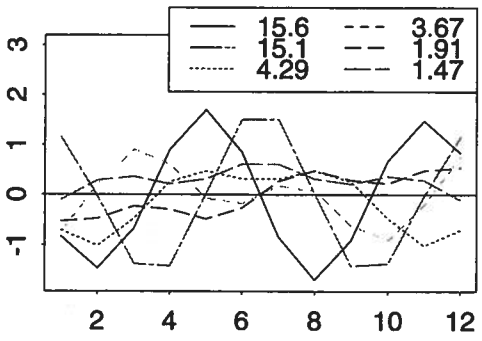

Figure 13.13: *First six time EOFs of an AR(2) process with* $\vec{a} = (0.9, -0.8)$ *obtained using window length* $m = 12$. *The patterns are normalized with the square root of the eigenvalue. The eigenvalues are given at the bottom.*

13.6.8 Multichannel Singular Spectrum Analysis. MSSA (see Vautard [380]) differs from SSA only in the dimension of the basic time series, which is now m'-dimensional rather than one-dimensional. The derived random vector $\vec{\mathbf{Y}}_t$ is therefore mm'-dimensional. Thus MSSA is *Extended EOF analysis* [13.1.8] in which m consecutively observed fields are concatenated together. The number of fields m is usually small compared with the field dimension m' in EEOF analysis. The opposite, $m > m'$, is often true in MSSA.

13.6.9 Estimation. We conclude with a brief comment on the estimation of eigenvalues and time EOFs in SSA. The same applies, by extension, to the eigenvalues and space-time EOFs in MSSA.

SSA is applied to a finite sample of observations $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ with $n \gg m$ by first forming $\vec{\mathbf{Y}}$-vectors,

$$\begin{aligned} \vec{\mathbf{y}}_1 &= (\mathbf{x}_1, \ldots, \mathbf{x}_m)^{\mathrm{T}} \\ \vec{\mathbf{y}}_2 &= (\mathbf{x}_2, \ldots, \mathbf{x}_{m+1})^{\mathrm{T}} \\ &\vdots \\ \vec{\mathbf{y}}_{n-m+1} &= (\mathbf{x}_{n-m+1}, \ldots, \mathbf{x}_n)^{\mathrm{T}}. \end{aligned}$$

Conventional EOF analysis is applied to the resulting sample of $n - m + 1$ $\vec{\mathbf{Y}}$-vectors. The estimated eigenvalues and EOFs can be computed from either the estimated covariance matrix of $\vec{\mathbf{Y}}$ (see [13.2.4]) or by means of SVD (see [13.2.8]).

Note that neither *North's Rule-of-Thumb* [13.3.5] nor Lawley's formulae (13.38, 13.39) can be used to assess the reliability of the estimate directly because consecutive realizations of $\vec{\mathbf{Y}}_t$ are auto-correlated (see (13.33)). The effects of temporal dependence must be accounted for (see Section 17.1) when using these tools.

Allen and co-workers [8, 9, 10] discuss the problem of discriminating between noisy components and truly oscillatory modes in detail.