Independent Component Analysis of Climate Data: A New Look at EOF Rotation

A. HANNACHI
Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

S. UNKEL AND N. T. TRENDAFILOV
School of Mathematics, Computing and Technology, The Open University, Milton Keynes, United Kingdom

I. T. JOLLIFFE
School of Engineering, Computing, and Mathematics, University of Exeter, Exeter, United Kingdom

(Manuscript received 9 April 2008, in final form 20 November 2008)

ABSTRACT

The complexity inherent in climate data makes it necessary to introduce more than one statistical tool to the researcher to gain insight into the climate system. Empirical orthogonal function (EOF) analysis is one of the most widely used methods to analyze weather/climate modes of variability and to reduce the dimensionality of the system. Simple structure rotation of EOFs can enhance interpretability of the obtained patterns but cannot provide anything more than temporal uncorrelatedness. In this paper, an alternative rotation method based on independent component analysis (ICA) is considered. The ICA is viewed here as a method of EOF rotation. Starting from an initial EOF solution rather than rotating the loadings toward simplicity, ICA seeks a rotation matrix that maximizes the independence between the components in the time domain. If the underlying climate signals have an independent forcing, one can expect to find loadings with interpretable patterns whose time coefficients have properties that go beyond simple noncorrelation observed in EOFs. The methodology is presented and an application to monthly means sea level pressure (SLP) field is discussed. Among the rotated (to independence) EOFs, the North Atlantic Oscillation (NAO) pattern, an Arctic Oscillation–like pattern, and a Scandinavian-like pattern have been identified. There is the suggestion that the NAO is an intrinsic mode of variability independent of the Pacific.

1. Introduction

Climate is a natural system that is characterized by complex and high-dimensional phenomena. Climate variations are the result of complex nonlinear interactions between many degrees of freedom. To gain insight in understanding the dynamical/physical behavior of the climate system, it is useful to attempt to understand their interactions in terms of a much smaller number of prominent modes of variability. This has led to the development of methods in the atmospheric science that give both a space and a time display of large space–time atmospheric data.

Empirical orthogonal function (EOF) analysis, also known as principal component (PC) analysis (PCA) (Jolliffe 2002), is among the most widely used methods in atmospheric science (Obukhov 1947, 1960; Fukuoka 1951; Lorenz 1956). Given any space–time meteorological field, EOF analysis finds a set of orthogonal spatial patterns, known as EOFs, along with a set of associated uncorrelated time series of PCs, such that the first few PCs account for as much as possible of the variation in the original field.

These geometric constraints characterizing EOFs and PCs can be very useful in practice because the covariance matrix of any subset of retained PCs is always diagonal. However, the constraints yield partially predictable
relationships between an EOF and the previous ones. The orthogonality constraint also makes the EOFs domain dependent and can be too nonlocal (Horel 1981; Richman 1986, 1987). These problems can cause difficulties in interpreting the obtained patterns (Aires et al. 2002; Dommenget and Latif 2002; Hannachi et al. 2007; Hannachi 2007a).

Various methods have been developed to overcome some of these difficulties. Linear transformations of EOFs based on rotation yield the concept of rotated empirical orthogonal functions (REOFs) (Horel 1981; Richman 1981, 1986). The method of rotation emerged in factor analysis and was motivated by both solving the indeterminacy problem in factor analysis and facilitating the factors’ interpretation (Browne 2001). The conventional REOFs have simpler and more localized structures compared to EOFs by rotating the vectors of loadings, hence losing some of the useful geometric properties of EOFs in favor of yielding better interpretation. An alternative to rotation and simplification was presented by Van den Dool et al. (2000) based on empirical orthogonal teleconnection (EOT). Van den Dool et al. (2000) argue that EOT can help the physical interpretation of the patterns. [For a comprehensive review of EOFs and REOFs, refer to Hannachi et al. (2007) and references therein.] Because REOFs cannot achieve anything more than temporal uncorrelatedness, it is desirable to find a rotation method that is able to enhance the physical interpretation of initial EOF patterns and goes one step further than conventional REOFs.

Recently, independent component (IC) analysis (ICA) has emerged as a strong competitor to PCA as an exploratory tool for signal data analysis (Hastie et al. 2001). Like PCA, ICA is a technique for separating mixtures of signals into their sources. Originally developed in the neural computation and signal-processing communities, this rapidly evolving method is currently finding applications in various disciplines, such as feature extraction, brain imaging, and financial econometrics (Hyvärinen and Oja 2000). Whereas there is a nonuniqueness problem in PCA that allows the arbitrary choice of rotation of axes without affecting the quality of the fit of the model, there is no rotational indeterminacy in ICA and hence no identifiability problem. This is achieved by looking for components that are both mutually statistically independent and nonnormal. Independence is a much stronger assumption than uncorrelatedness. Independence implies uncorrelatedness, whereas the converse only holds true when the variables involved are jointly multivariate Gaussian. The concepts developed here make use of three different, but interrelated notions of random quantities, namely, uncorrelatedness, orthogonality, and independence.1

In fact, the ICA can be considered as an EOF rotation method. Starting from an initial EOF solution, rather than rotating the loadings toward simplicity, ICA seeks a rotation matrix that maximizes the independence between the components. The rotational redundancy of PCA is removed by using supplementary higher-order information for nonnormal variables not contained in the sample covariance matrix.

ICA is based on the assumption that if different sources stem from different physical processes, then those sources are statistically mutually independent. Accordingly, ICA separates signal mixtures into independent sources and, if independent components can be found, they are identified with the hidden sources. The ICA can be of interest in climate research to separate underlying anomaly signals that may have an independent forcing. If one finds statistically independent patterns of variability in the climate system, then it is reasonable to assume that those patterns are generated by quasi-independent physical processes. Perhaps a good example is the ongoing debate within the climate community on whether the North Atlantic Oscillation (NAO) and the North Pacific Oscillation (NPO) are two different faces of the same coin [i.e., the Arctic Oscillation (AO)] (Ambaum et al. 2001; Vallis et al. 2004; Dommenget 2007).

The use of ICA in climate research is quite recent and the number of research papers is limited. Philippon et al. (2007) employ ICA to extract independent modes of interannual and intraseasonal variability of the West African vegetation. Mori et al. (2006) applied ICA to monthly sea level pressures (SLPs) to find the main independent contributors to the AO signal. [See also Basak et al. (2004) for an analysis of the NAO, Fodor and Kamath (2003) for an application of ICA to global temperature series, and Aires et al. (2000) for an analysis of tropical sea surface temperatures.] The ICA technique has the potential to avoid the PCA “mixing problem.” PCA has the tendency to mix several modes of comparable magnitude, often generating spurious regional overlaps or teleconnections where none exists.

---

1 These terms are nicely compared within both an algebraic and geometric framework in a didactic paper by Rodgers et al. (1984), and we give here a brief description of these concepts. First, we note that independence and orthogonality refer to noncentered variables, whereas uncorrelatedness refers to centered variables. Two variables are linearly dependent when there is a (nontrivial) linear relationship between them. Orthogonality, on the other hand, means perpendicularity, whereas uncorrelatedness means orthogonality of the centered variables. Orthogonality is a particular case of independence.
or distorting existing overlaps or teleconnections (Aires et al. 2002).

The purpose of the present paper is to show the usefulness of ICA in climate research as a method of EOF rotation and provide a tool to the atmospheric scientist to help analyze and interpret climate patterns. Aires et al. (2002) also provide a discussion of ICA in connection with EOF rotation. Their procedure is based on a neural network algorithm and is rather complicated. The approach presented here is much more straightforward. By minimizing a criterion based on the sum of squared fourth-order statistics, the initial principal components are rotated toward independence. In practice, the first leading EOFs and PCs can easily be computed using the singular value decomposition (SVD) (Golub and van Loan 1996). Assume that \( X \) has rank \( r \) with \( r \leq \min(n,p) \). Using the SVD, \( X \) can be decomposed as

\[
X = \mathbf{M} \mathbf{U}' = \sum_{j=1}^{r} \lambda_j \mathbf{m}_j \mathbf{u}_j',
\]

where \( \mathbf{M} = (\mathbf{m}_1, \ldots, \mathbf{m}_r) \in \mathbb{R}^{n \times r} \) and \( \mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_r) \in \mathbb{R}^{p \times r} \) are orthonormal matrices so that \( \mathbf{M}' \mathbf{M} = \mathbf{U}' \mathbf{U} = \mathbf{I}_r \), with \( \mathbf{I}_r \) being an identity matrix of order \( r \), and \( \Lambda = \mathbf{M} \mathbf{L} \mathbf{M}' \) is a diagonal matrix with the singular values of \( X \), sorted in decreasing order, \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r \geq 0 \), on its main diagonal. The prime notation stands for the transpose operator. The matrix \( \mathbf{U} \) in (1) is the matrix of EOFs and \( \mathbf{M} \) is the matrix of PC scores. The variance of the \( j \)th PC, \( j = 1, \ldots, r \), is \( \lambda_j / (n-1) \), which is equal to the \( j \)th eigenvalue of the sample covariance matrix of \( X \). In practice, the first leading \( k \) components, with \( k < r \), account for a substantial proportion of the total variance in the data (e.g., 80%) and the sum in (1) is therefore truncated after the first \( k \) terms. So an EOF analysis comes down to finding a \( (n \times k) \) real matrix of component scores of the \( n \) time samples on the \( k \) components and a loading matrix \( \mathbf{U} \in \mathbb{R}^{p \times k} \) containing the EOF coefficients for the \( p \) grid points on the \( k \) components.

b. Rotated EOFs

The principal objective of (conventional) EOF rotation is to obtain patterns having a more localized simple structure that could ease their physical interpretation and thereby overcome some of the drawbacks of EOFs mentioned above. Rotation acts on a prescribed subset of EOFs and can be performed either in an orthogonal or oblique fashion. Several possible analytic orthogonal and oblique rotation criteria exist in the literature (Harman 1976; Browne 2001). To aid interpretation, all criteria are designed to make the EOF coefficients as simple as possible in some sense, with most loadings to have values either close to or far from zero and as few as possible intermediate values.

Given an initial \( (p \times k) \) loading matrix \( \mathbf{U}_k = (\mathbf{u}_1, \ldots, \mathbf{u}_k) \) of the leading \( k \) EOFs, an orthogonal rotation is formally achieved by seeking a matrix \( \mathbf{T} \in \mathbb{R}^{k \times k} \) satisfying \( \mathbf{T}' \mathbf{T} = \mathbf{T} \mathbf{T}' = \mathbf{I}_k \), such that the rotated loadings

\[
\mathbf{B} = \mathbf{U}_k \mathbf{T}
\]

optimize a specific simplicity criterion \( f(\mathbf{B}) \). For orthogonal rotation, the Orthomax procedure maximizes
where $b_{ij} (i = 1, \ldots, p; j = 1, \ldots, k)$ are the elements of $B$. By choosing different values of $\gamma$, the objective function (3) provides a family of criteria for simple structure rotation. For example, $\gamma = 1$ yields the (raw) Varimax criterion. The idea of Varimax is to maximize the variance of the squared loadings within each column of $B$ and drive the squared loadings toward the end of the range $[0, 1]$, and hence the loadings toward $-1$, 0, or 1 and away from intermediate values, as required. A normalized version of the Varimax is also often used in which the elements $b_{ij}$ in (3) are replaced by $b_{ij}/\sqrt{\sum_{j=1}^{k} b_{ij}^2}$.

For $\gamma = 0$, the function (3) corresponds to the Quartimax criterion, which attempts to maximize the sum of the fourth powers of the loadings. In general, the structures obtained using Quartimax tend to be less local than those obtained using Varimax (Hannachi et al. 2006). In addition, Richman (1986) argues that Varimax is slightly less sensitive to changes in size of the spatial domain than Quartimax. As suggested by Browne (2001), the minimum entropy-based rotation in general leads to a better simple structure than Varimax. Such EOF rotation has not been tried in climate applications.

With the usual normalization $U_kU_k^T = I_k$, the rotated EOFs remain orthogonal but the components are correlated. Besides rotating $U_k$, one could alternatively rotate the scaled EOFs $U_k \Lambda_k$, where $\Lambda_k = \text{diag}(\lambda_1, \ldots, \lambda_k)$ represents the diagonal matrix containing the leading $k$ singular values of $X$. Because the columns of $U_k$ have unit length, the matrix $U_k \Lambda_k$ has columns that have lengths equal to the corresponding singular values. For this normalization, which is often used in atmospheric science, the rotated PCs are correlated and have loadings that are not orthogonal. The reason for this is based on arguments taken from factor analysis (Jolliffe 2002, p. 272). To obtain uncorrelated components, a different scaling $U_k \Lambda_k^{-1}$ can be applied but the loadings are again not orthogonal (Jolliffe 2002). Hence, either one or both properties possessed by EOF analysis are lost by an EOF rotation.

Unlike orthogonal rotation, oblique rotation methods seek a nonorthogonal and nonsingular rotation matrix $T \in \mathbb{R}^{p \times k}$ with columns having unit length, such that the oblique rotated loadings minimize a particular criterion such as the Geomin (Browne 2001). Oblique rotations give extra flexibility and often produce a better simple structure than orthogonal rotations. Of course, with oblique rotations, neither of the two properties possessed by EOF analysis is retained.

In the next section, the method of ICA is described that can be considered a specific method of EOF rotation in PCA or EOF analysis. Rather than rotating the EOFs toward simplicity, the component scores are rotated orthogonally toward independence.

### 3. Independent component analysis

#### a. Statistical model and identifiability

Consider the following linear latent variable model in which all variables are assumed to be measured at least on an interval scale:

$$x = As,$$  \hspace{1cm} (5)

where $x \in \mathbb{R}^{p \times 1}$ is a random vector of manifest variables; $s \in \mathbb{R}^{k \times 1}$ is a random vector of $k \leq p$ latent variables called sources, signals, or components; and $A \in \mathbb{R}^{p \times k}$ is a mixing matrix of fixed coefficients. The mixing matrix $A$ is required to have full column rank. Assume that $E(x) = 0$ and $E(s) = 0$, where $E(\cdot)$ is the expectation operator. Furthermore, suppose that $s$ consists of mutually independent sources of which at most one is Gaussian and whose densities are square integrable. Using these assumptions, Eq. (5) corresponds to the (noise free) ICA model (Lee 1998).

The classic example for which ICA is useful is the so-called cocktail-party problem. Imagine that the guests of a party in a large room are clustered into conversational groups and several people are assumed to speak simultaneously. Suppose that someone wants to eavesdrop on these conversations by placing microphones in the corners of the room. The output of each microphone is a mixture of the voice signals emitted by the speakers. Given these signal mixtures, ICA can recover the original voices or sources. Although this example uses speech, ICA can extract sources from any set of two or more measured signal mixtures. Whereas the extension of this analogy to climate is certainly challenging because of the presence of nonlinearity, the point is that this nonlinearity is manifest mainly on short (e.g., synoptic or baroclinic) time scales and that on longer time scales the system is nearly quasi-linear. For short time scales, ICA can still be used, however, as an exploratory tool in a manner similar to that of EOF analysis.
Given a multivariate sample of \(n\) independent observations on \(\mathbf{x}\), the ICA model (5) can be written as
\[
\mathbf{X} = \mathbf{S}\mathbf{A}^* ,
\]
where \(\mathbf{X} \in \mathbb{R}^{n \times p}\) is the observed data matrix and \(\mathbf{S} \in \mathbb{R}^{n \times k}\) denotes the unknown matrix of scores of the \(k\) independent sources on \(n\) observations. The ICA problem is to recover \(\mathbf{S}\) from \(\mathbf{X}\) without knowing \(\mathbf{A}\). Without loss of generality, it can be assumed that the independent sources are standardized to unit variance. Then, the mixing matrix and hence the latent signals can be identified up to trivial ambiguities in sign and order (Hyvärinen et al. 2001). In other words, a demixing matrix \(\Gamma^\prime \in \mathbb{R}^{k \times p}\) is sought such that
\[
\Gamma^\prime \mathbf{A} = \Delta ,
\]
where \(\Delta \in \mathbb{R}^{k \times k}\) is a signed permutation matrix, that is, \(\Delta = \mathbf{PD}\), where \(\mathbf{P} \in \mathbb{R}^{k \times k}\) is a permutation matrix that has a single entry of 1 in each row and column and \(\mathbf{D} \in \mathbb{R}^{k \times k}\) is a diagonal matrix with entries \(d_{ij}\) satisfying \(|d_{ij}| = 1\) for \(i = 1, \ldots, k\).

\[\text{b. Data preprocessing and dimensionality reduction by PCA}\]

To find \(\Gamma\), the first step is to apply a preprocessing technique in which \(\mathbf{X}\) is transformed, using a whitening or sphering operation, into a new data matrix \(\mathbf{Y} \in \mathbb{R}^{n \times k}\), with \(\mathbf{I}_k\) as its covariance matrix.

Suppose the aim is to reduce the dimensionality of the data to \(k\) dimensions. Then the whitening of the data matrix \(\mathbf{X}\) is obtained through PCA and yields
\[
\mathbf{Y} = \mathbf{XU}_k\Lambda_k^{-1} = \mathbf{XW}^\prime ,
\]
where \(\Lambda_k^{-1} = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_k^{-1})\) is a diagonal matrix containing the inverses of the leading singular values of \(\mathbf{X}\), and \(\mathbf{U}_k = [\mathbf{u}_1, \ldots, \mathbf{u}_k]\) is the (orthogonal) matrix of the associated \(k\) leading EOFs.

ICA attempts to go one step further than PCA by finding a matrix \(\mathbf{T}\) that transforms the sphered data \(\mathbf{Y}\) into
\[
\hat{\mathbf{S}} = \mathbf{Y}\mathbf{T} = \mathbf{XW}'\mathbf{T} ,
\]
where \(\hat{\mathbf{S}}\) is an estimate of \(\mathbf{S}\) in (6). Hence, sphering via PCA determines the sources up to an orthogonal transformation. In other words, referring to (7), after sphering the ICA search for the demixing matrix \(\Gamma\) boils down to seeking an orthogonal matrix \(\mathbf{T}\) such that
\[
\mathbf{T}'\mathbf{W}\mathbf{A} = \Delta .
\]
Using only second-order statistics, PCA only decorrelates the data and the latent sources can only be estimated up to an orthogonal rotation. Hence, PCA is not able to separate linear mixtures into their independent sources (the identifiability problem).

There are two fundamental estimation principles that enable ICA to perform source separation. The two principles, described briefly in appendix A, are nonlinear decorrelation and maximizing nonnormality (Hyvärinen et al. 2001). Regarding the last principle, which is most relevant here, it is known that nonnormality is a big issue in climate. Non-Gaussian behavior originates from various sources, such as the effect of decorrelating nonlinearity resulting in multiplicative noise (Sura et al. 2005), asymmetric tropical forcing (e.g., El Niño), saturation at the extremes, and nonlinear regime behavior (see, e.g., Hannachi 2007b for more discussion and further references). Non-Gaussian behavior is manifest particularly at short time scales (Holzer 1996). Time averaging, taken particularly over long periods, is expected to lead to Gaussian behavior. However, a main feature of climate is the presence of strong autocorrelations even over relatively long lags. This means that time-averaged fields will still inherit some of its characteristics (e.g., nonnormality).

There is a wide class of ICA algorithms that achieve approximate independence by optimizing criteria involving higher-order cumulants; for example, the Joint Approximative Diagonalization Eigenmatrices (JADE) criterion proposed by Cardoso and Souloumiac (1993) performs joint diagonalization of a set of fourth-order cumulant matrices. The Orthomax-based criteria proposed in Kano et al. (2003) are quadratic and linear functions of fourth-order statistics. Unlike higher-order cumulant-based methods, the popular FastICA algorithm chooses a single nonquadratic smooth function [e.g., \(g(x) = \log(\cosh(x))\)] such that the expectations of this function yield a robust approximation to negentropy (Hyvärinen et al. 2001). In the next section, a criterion is introduced that requires the minimization of the sum of squared fourth-order statistics formed by covariances computed from squared components.

4. Rotation toward independence

\[\text{a. Rotation criterion and optimization algorithm}\]

Finding a matrix of uncorrelated component scores \(\mathbf{Y}\) and a matrix of EOFs \(\mathbf{U}\) is a standard problem in EOF analysis. To solve the corresponding ICA problem in (6), one needs to rotate \(\mathbf{Y}\) toward independence, that is,
\[
\mathbf{S} = \mathbf{Y}\mathbf{T} ,
\]
for some orthogonal matrix \(\mathbf{T}\). To find the matrix \(\mathbf{T}\) that leads to approximately independent component scores,
an appropriate rotation criterion is set up next (Jennrich and Trendafilov 2005).

Recall that if the components \( s_1, \ldots, s_k \) are independent their squares \( s_1^2, \ldots, s_k^2 \) are also independent. Thus, the model covariance matrix of the squared components is diagonal. Now let \( \mathbf{V} \) be an arbitrary orthogonal matrix and let

\[
\mathbf{G} = \mathbf{YV}.
\]  

(12)

The sample covariance matrix between the element-wise squares of \( \mathbf{G} \) is

\[
\mathbf{R} = \frac{1}{n-1} (\mathbf{G} \odot \mathbf{G})' \mathbf{H} (\mathbf{G} \odot \mathbf{G}),
\]  

(13)

where \( \mathbf{H} = [\mathbf{I}_n - (1/n)\mathbf{1}_n \mathbf{1}_n'] \) is the centering matrix, \( \mathbf{1}_n = (1, \ldots, 1)' \), and \( \odot \) denotes the element-wise (Hadamard) matrix product.

Consider the following rotation criterion to be minimized:

\[
\mathcal{F}(\mathbf{V}) = \frac{1}{2} \left( \| \mathbf{R} \|_F - \| \text{diag}(\mathbf{R}) \|_F \right),
\]  

(14)

where \( \| \mathbf{R} \|_F = \text{trace}(\mathbf{R} \mathbf{R}') \) is the Frobenius norm of \( \mathbf{R} \). Hence, the criterion (14) minimizes the sum of the squared off-diagonal elements of \( \mathbf{R} \) over all orthogonal matrices \( \mathbf{V} \) or equivalently over all orthogonal rotations \( \mathbf{V} \) of \( \mathbf{Y} \). It is now clear that for model (11) (in data space), criterion (14) satisfies \( \mathcal{F}(\mathbf{T}) = 0 \), because the model covariance matrix \( s_1^2, \ldots, s_k^2 \) is diagonal. Therefore, the matrix \( \mathbf{T} \) may be found among the minima of \( \mathcal{F} \) using the gradient projection algorithm (see appendix B for details).

b. An illustrative example

The procedure described above shall be illustrated with a simple example by considering two hypothetical models of the AO pattern (Ambaum et al. 2001; Itoh 2002; Mori et al. 2006). Let \( \mathbf{x} = (x_1, x_2, x_3)' \) denote a vector of observed time series representing the sea level pressure anomalies at the centers of action of the AO pattern located in the Arctic, North Pacific, and North Atlantic, respectively. Furthermore, \( \mathbf{s} = (s_1, s_2)' \) with \( s_1 \) and \( s_2 \) being mutually independent time series that correspond to some underlying source signals. In the two models, the observed variables are considered as linear combinations of the latent signals as follows:

\[
\text{AO-Model 1: } \mathbf{x} = \mathbf{A}_1 \mathbf{s} \quad \text{and} \quad \text{AO-Model 2: } \mathbf{x} = \mathbf{A}_2 \mathbf{s},
\]  

(15) (16)

where the respective mixing matrices \( \mathbf{A}_1 \) and \( \mathbf{A}_2 \) are given by

\[
\mathbf{A}_1 = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_2 = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}.
\]  

Assume that \( s_1 \) and \( s_2 \) are uniformly distributed on the interval \([-1, 1]\). Then, for models (15) and (16), the covariance matrices of \( \mathbf{x} \) are

\[
\mathbf{A}_1 \mathbf{A}_1^T = 2 \mathbf{A}_2 \mathbf{A}_2^T,
\]  

(17)

hence, a conventional EOF analysis is unable to distinguish the intrinsic structure between the two models.

However, if the component scores are plotted within the EOF phase space, the two models are distinguishable. Given a random sample of 500 independent realizations for \( s_1 \) and \( s_2 \), the scores of the two leading EOFs are plotted within the EOF phase space and are shown in Fig. 1a for AO-Model 1 and in Fig. 1b for AO-Model 2. The joint density of the latent signals \( s_1 \) and \( s_2 \) is uniform on a square. In Fig. 1a, the joint density of the scores of EOF 1 and EOF 2 for AO-Model 1 is uniform on a square and PC 1 and PC 2 are independent. In Fig. 1b, the joint density of the scores of EOF 1 and EOF 2 cannot be expressed as the product of the marginal densities and the components are not independent.

If the ICA rotation procedure, as described in the previous section, is applied instead, the independent components are extracted successfully\(^3\) for both cases

\(^3\) Note that, for example, the Varimax rotation procedure is not helpful in this regard because it is based on simplifying the structure of the EOFs, which are identical in this example [Eq. (17)], and for both examples the EOFs are identical to the rotated EOFs.
Therefore, unlike an EOF analysis, the ICA is able to reveal the intrinsic structure embedded in the data for both models.

5. Application to monthly sea level pressure

a. Data description

The proposed approach is applied to data from the NCEP–NCAR reanalysis project (Kalnay et al. 1996; Kistler et al. 2001). The dataset consists of monthly mean SLPs over the Northern Hemisphere (NH) north of 20°N, is available on a 2.5° × 2.5° regular grid and spans the period January 1948–December 2006. Prior to the analysis, the data were preprocessed as follows: First, the mean annual cycle was calculated by averaging the monthly data over the years. Anomalies were then computed as departures from the mean annual cycle. Finally, an area weighting was performed by multiplying the SLP anomalies by the square root of the cosine of the corresponding latitude. The area weighting is applied to reduce the effect of high-latitude data that correspond to smaller grid sizes and hence are subject to greater sampling variation4 (Baldwin et al. 2009). These weighted SLP anomalies are considered in the following. The whole-year monthly data are used only to have a reasonable sample size and also to compute correlations (later) with sea surface temperature and global SLP. Few comments, however, will be mentioned later about the winter monthly data.

b. Results

We have tested the data for normality using a Kolmogorov–Smirnov (KS) test. Figure 2 shows the grid points where the null hypothesis of normality is rejected at 1% significance level. Clearly, most of the NH SLP anomalies depart from normality. As outlined previously, some of this nonnormality is due to the high autocorrelation at shorter time scales (see, e.g., Madden 1976). Blocking phenomena, which can last for several weeks, can also lead, via skewness, to nonnormality (Barnston and Van den Dool 1993a,b), in addition to other asymmetric forcing (e.g., from the tropical Pacific Ocean). Dommenget (2007) also tested the Northern Hemisphere SLP data against a null hypothesis of a first-order spatial autoregressive model and found that the leading EOFs depart significantly from those of the null hypothesis.

From the weighted anomaly data matrix, the sample covariance matrix was computed first. Figure 3 shows the spectrum of the sample covariance matrix expressed in percent of explained variance of the EOFs, along with the approximate 95% confidence limits given by the rule of thumb of North et al. (1982). An effective sample size of 180 has been used, based on the estimate of Thiébaux and Zwiers (1984) using the lag-1 autocorrelation function of the first component. We have restricted our analysis to the leading $k = 5$ EOFs, explaining around 52% of the total variance in the data. Figure 4 displays the leading two EOFs showing the familiar AO and the Pacific pattern (see also Hannachi et al. 2006, 2007).

The PCs corresponding to the leading $k = 5$ EOFs were then rotated toward independence by optimizing the objective function (14) using the gradient projection algorithm and many initial conditions. The final cost function is on the order of $O(10^{-16})$ with a CPU time of $O(10^4)$ seconds on a Sun workstation. The REOFs are obtained by postmultiplying the matrix of scaled EOFs $\mathbf{U}_2 \mathbf{A}_2$ with the optimal orthogonal rotation matrix $\mathbf{T}$ that minimizes the criterion (14). The approximate ICs have no natural ordering. For convenience, they are ordered according to the variance of the projection of the SLP anomalies onto the rotated EOFs. So the first REOF and the associated IC have the largest variance compared to the remaining REOFs.

Figure 5 shows the percentage of explained variance associated with these REOFs along with their confidence intervals (North et al. 1982) using an effective

---

4 It is pointed to us by one referee that weighting is recommended only when an integral in space is involved.
sample size of 180. Note that for rotated components, North’s rule of thumb, which is formulated for EOFs, is simply used to get broad estimates of the confidence intervals. Note that various REOFs are degenerate in the sense that they explain a similar amount of variance but still have approximately independent associated components.

The first 5 ICs are shown in Fig. 6. Their cross correlations have been checked and found to be zero according to some prescribed level of machine precision. The cross correlations of various nonlinear transformations of ICs have also been computed and compared to those obtained using PCs. The upper triangular part of the matrix in Table 1 shows the absolute values of the correlations between the element-wise fourth power of ICs 1–5. The lower part shows the corresponding correlations using the PCs instead of ICs. Significant correlations at the 1% and 5% levels are indicated below the diagonal. Table 2 is similar to Table 1, but now the nonlinear transformation is the absolute value of the third power law. Note again the significant correlations for the transformed PCs in the lower triangular part of Table 2, whereas no such significant correlations are obtained with the ICs. We have repeated the same experiment using various other nonlinear functions, such as \( g(x) = \exp(x^3) \) and \( g(x) = \exp(|x|) \), and we found similar results to those shown in Tables 1 and 2. We note here that when we use the Varimax procedure, the time series associated with the REOFs (not shown) are well correlated and the correlation coefficients (in absolute value) range between 0.26 and 0.70.

A very useful exploratory tool to test the null hypothesis of normality is to use the quantile–quantile (q–q) plot, in which the quantiles of a given IC are plotted versus those of the standard normal distribution. Figure 7 shows the q–q plots of all SLP ICs. The straight diagonal lines in Fig. 7 are for the normal distribution and any departure from these lines reflects nonnormality. Clearly, all q–q plots display strong nonlinearity (i.e., nonnormality). A formal KS test reveals that the null hypothesis of normality is rejected for the first three ICs at 1% significance level and for the last two ICs at 5% level. The nonnormality of the PCs has also been checked and compared with that of the ICs using the q–q plot (not shown). It is found that the ICs are more nonnormal than the PCs.

The spatial patterns of the REOFs are shown in Fig. 8. As stated above, their order is fixed by the variance explained by the projection of the SLP anomalies onto them (see also Fig. 5). The first REOF (Fig. 8a) shows two main centers of action with opposite polarity over the north Eurasian border and North Pacific and a weaker center over the northwestern Atlantic. This pattern is reminiscent of the AO, particularly over the Pacific and polar regions. The second REOF (Fig. 8b) shows unambiguously the familiar NAO pattern with major centers located over Iceland and the northwest
Iberian Peninsula, respectively. The third rotated pattern (Fig. 8c) shows two centers located over south Scandinavia and the northwestern Pacific and an opposite-sign center over southern Greenland and northern Canada. The fourth pattern (Fig. 8d) looks like the Scandinavian pattern, but the north Russia center stretches far southeast to the Aleutian. The last one (Fig. 8e) has two main opposite polarity centers over the North Atlantic and North Pacific and a weaker center over Russia. We have also looked at the winter monthly REOFs. Three winter REOFs (not shown) look similar to REOFs 2, 3, and 5 (Figs. 8b,c,f) and the remaining two REOFs (not shown) look less similar to REOFs 1 and 4 (Figs. 8a,d). For example, one of these two REOFs (not shown) shows two centers of opposite polarity, one stretching from the northeast Atlantic, through the British Isles, to northern Russia, and to the second center over eastern Aleutian/Alaska. The other REOF (not shown) bears some similarities to REOF 4 (Fig. 8d) over the North Atlantic and Eurasia and the Pacific center has opposite polarity.

The REOF2 (Fig. 8b) has been compared to the observed NAO. We consider the NAO index taken from the Climate and Global Dynamic Division of NCAR (available online at http://www.cgd.ucar.edu/cas/jhurrell/indices.html), which is a monthly station-based index obtained from the difference of normalized SLP between Ponta Delgata, Azores, and Stýkkisholmur/Reykjavik, Iceland. The correlation coefficient between the NAO index and IC2 component is then computed and is found to be about 0.63. A similar comparison has been performed between REOF5 and the Pacific pattern. The latter is obtained through a Varimax rotation (Hannachi et al. 2007) of the SLP EOFs. The correlation coefficient between the obtained Pacific Oscillation index and IC5 is about 0.67. The spatial correlation between REOF2 (Fig. 8b) and the Varimax NAO pattern is 0.9. Here, REOF1 (Fig. 8a), which is more related to the AO pattern, has about 0.84 spatial correlation with EOF1 (Fig. 4a). Most of this correlation comes from the Pacific and the polar centers. Over the North Atlantic region, REOF1 (Fig. 8a) and REOF2 (Fig. 8b) are nearly orthogonal. This suggests that the NAO is an intrinsic mode of variability independent of the Pacific, and that the AO can be regarded as being mostly controlled by the Pacific center.

Global teleconnection of the independent components has been investigated. Using the global monthly SLP field, only IC4 is found to have tropical connection, particularly from the western tropical Pacific. Figure 9 shows the correlation between IC4 and the global monthly SLP, where only significant correlations at 1% level are plotted. Teleconnections with sea surface temperature (SST) have also been computed. We have used the Hadley Centre Ice and Sea Surface Temperature (HadISST; Folland et al. 1999) monthly SST for the period of January 1948–November 2006. The largest correlations are obtained with IC1 and IC2. Figure 10 shows the correlation map with the leading component IC1 where all shown correlations are significant at 1% level. We recognize the North Pacific Oscillation pattern. There are also weaker but significant correlations with the Niño-3 region and the northwestern tropical Atlantic: the strongest correlations are about -0.37 and 0.28, respectively. The same correlation map but for IC2 is shown in Fig. 11, in which the strongest correlations are about -0.36 and 0.34. We recognize here the classical picture of the North Atlantic tripole pattern associated with NAO. These correlation maps help interpret the obtained independent components. In fact, these maps correspond to the SLP ICs leading the SSTs by one month.

6. Discussion and conclusions

Rotated EOFs have been introduced to overcome some of the drawbacks related to the orthogonality and uncorrelatedness of EOFs and PCs, respectively, and also to enhance interpretation of the obtained patterns. Conventional simple structure rotation methods originated in factor analysis attempt to rotate a fixed number of EOF patterns in either an orthogonal or an oblique fashion to optimize a certain simplicity criterion. In particular, orthogonal rotation has been widely used to
analyze weather and climate modes of variability. However, by rotating EOFs toward a simple structure, the properties possessed by EOFs are lost.

Unlike an EOF analysis, ICA not only decorrelates the data, but it goes one step further than PCA by finding sources that are statistically mutually independent. In addition, at most one component is allowed to be normally distributed. The assumption of nonnormal and independent components removes the rotational indeterminacy of PCA. Up to trivial ambiguities, the ICA solution is unique and hence there is no identifiability problem. In fact, the ICA can be considered as a specific rotation method in EOF analysis. Starting from an initial EOF solution, rather than rotating the loadings toward simplicity, the components were rotated orthogonally toward independence. The ICA is based on the assumption that if different sources stem from different physical processes, then those sources are statistically mutually independent. Hence, the ICA can be very useful in climate studies to separate underlying anomaly signals that may have an independent forcing.

In this paper, the ICA rotation was applied to the monthly NCEP–NCAR SLP field for the period of January 1948–November 2006 over the Northern Hemisphere. The ICs have been found to be approximately independent by computing correlations between various nonlinear transformations of the rotated components. It was shown that the ICA rotation can enhance the interpretability of the prominent sea level pressure EOFs. Among the spatial REOFs (i.e., the spatial patterns associated with the ICs), we found the NAO pattern, an AO-like pattern, and a Scandinavian-like pattern to be approximately independent modes of variability. There is suggestion that the NAO is an intrinsic mode independent of the Pacific and that the AO can be regarded as being controlled mainly by the Aleutian low. One rotated EOF is found to have tropical teleconnection. Significant 1-month lead correlations are also identified between some ICs and monthly means SSTs.

**TABLE 1. Correlation matrix of the fourth power elements of ICs 1–5 (above the diagonal) and the same correlation but for the PCs (below the diagonal). The sign of the correlations has been dropped.**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.08*</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0.01</th>
<th>0.02</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08*</td>
<td>1</td>
<td>0.01</td>
<td>1</td>
<td>0</td>
<td>0.010</td>
<td>0.01</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.01</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>0.08*</td>
<td>1</td>
<td>0.02</td>
<td>0.01</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1**</td>
<td>0.05</td>
<td>0.04</td>
<td>0.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant correlation at 5% level.
** Significant correlation at 1% level.

**TABLE 2. As in Table 1, but using third power of absolute value function.**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.02</th>
<th>0</th>
<th>0.02</th>
<th>0.03</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.13*</td>
<td>1</td>
<td>0.04</td>
<td>0.05</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.05</td>
<td>1</td>
<td>0.03</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.04</td>
<td>0.11</td>
<td>1</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>0.14*</td>
<td>0.09**</td>
<td>0.07**</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

* Significant correlation at 1% level.
** Significant correlation at 5% level.
The rotation criterion requires minimization of squared fourth-order statistics formed by covariances computed from squared components. The criterion used was chosen primarily because its appropriateness was motivated in a straightforward fashion. It is easily optimized using the gradient projection algorithm. The same results were obtained using an alternative orthogonal rotation algorithm employing ordinary differential equations (Trendafilov 2006). A number of other criteria for rotating the components toward independence have also been tried (e.g., the optimizing the generalized or total variance of the covariance matrix of the squared components), but the authors will not report further on this here.

The ICA methodology via EOF rotation has been applied here using five EOFs and five PCs. Sensitivity to truncation is, however, an important issue in addition to whether a given truncation is adequate to represent the unknown independent components; we note here that the mixing matrix was assumed of full rank. These are topics for further investigation. For example, the issue of truncation is being investigated under a more general framework of noisy ICA where model (6) is allowed to include a noise term. The criterion used here is based on fourth-order statistics and may not be robust against outliers. The criterion is used because of its smooth properties useful for optimization. However, other robust fourth-order statistics could be used and this is also a topic for further investigation.

Unlike conventional rotation, the rotation toward independence yields patterns that may not be simple. This is because the scope of the present approach is
totally different from that of conventional rotations. Conventional rotations are based on simple geometry, whereas rotation criteria based on independence are much deeper. The rotated EOFs obtained in this manuscript show structures over most of the domain. It is possible to attempt a rotation criterion that simultaneously achieves approximate (temporal) independence and also simplicity of the (spatial) patterns, which may enhance further the physical interpretability of the REOFs. This point is beyond the scope of this paper and is left for future research.

Acknowledgments. Some of this work was performed while the first author was visiting the Walker Institute in the Department of Meteorology, Reading University. The authors thank two anonymous reviewers for their constructive and challenging comments that helped improve the manuscript.

APPENDIX A

ICA Estimation Principles

a. Nonlinear decorrelation and maximum nonnormality

As in section 3, we let \( \mathbf{s} \in \mathbb{R}^{k \times 1} \) be a random vector of \( k \leq p \) latent variables \( s_1, \ldots, s_k \). Independent latent sources \( s_1, \ldots, s_k \) satisfy

\[
E[g(s_i)h(s_j)] = E[g(s_i)]E[h(s_j)] \quad \text{for} \quad i \neq j,
\]

where \( g(s_i) \) and \( h(s_i) \) are any absolutely integrable functions of \( s_i \) and \( s_j \), respectively. Hence, independence implies nonlinear uncorrelatedness. Thus, one could attempt to implement ICA by a stronger form of decorrelation where the estimated components \( \hat{s}_1, \ldots, \hat{s}_k \) are uncorrelated even after some nonlinear transformation, that is, the correlation of the transformed recovered components should satisfy

\[
\rho[g(\hat{s}_i), h(\hat{s}_j)] = 0 \quad \text{for} \quad i \neq j, \quad (A1)
\]

where \( g(\cdot) \) and \( h(\cdot) \) are some suitable nonlinear functions. If the nonlinearities are chosen properly, the independent components can be recovered approximately (Hyvärinen et al. 2001).

Another estimation principle is maximization of non-Gaussianity (Hyvärinen et al. 2001, section 8.1). This principle is based on Lyapunov’s central limit theorem (CLT) (Pawitan 2001). According to the CLT, a sum of two independent nonnormal random variables is closer to the normal distribution than the original ones. Take a linear combination of the observed mixture variables, which in turn is also a linear combination of the independent components. By the CLT, a linear combination of two or more independent sources will be maximally nonnormal if it equals one of the independent components. Estimating \( s_1, \ldots, s_k \) can therefore be accomplished by finding the right linear combination of the mixture variables that maximizes its nonnormality.
Note the close connection between ICA and exploratory projection pursuit devised by Friedman and Tukey (1974) and Friedman (1987). The latter technique can reveal interesting structures through orthogonal projection onto selected low-dimensional state spaces. It has been argued that the normal distribution is the least interesting one and that the most interesting projections are those that show the least normal distribution. Whereas projection pursuit looks merely for nonnormal projections of the data, ICA seeks for both nonnormal and independent components.

For normally distributed data, the covariance or correlation matrix is a sufficient statistic for the multivariate normal distribution and so higher-order cross products do not add any information (Mooijaart 1985). In ICA, the latent signals and hence the data are not independent and the joint density factorizes. Hence, source separation can be evaluated by the following contrast: \( \phi_{\text{MI}}[\mathbf{s}] = I[\mathbf{s}] \), where \( \phi_{\text{MI}}[\mathbf{s}] \) is minimized when the separation into independent sources has been successful, that is, if \( \phi_{\text{MI}}[\mathbf{s}] = \phi_{\text{MI}}[\mathbf{s}] = 0 \).

Alternatively, mutual information may be interpreted as a distance by using the Kullback–Leibler (KL) divergence. This is defined between two pdfs \( f_\mathbf{x}(\mathbf{x}) \) and \( g_\mathbf{x}(\mathbf{x}) \) as

\[
D_{\text{KL}}[f_\mathbf{x}(\mathbf{x}), g_\mathbf{x}(\mathbf{x})] = \int f_\mathbf{x}(\mathbf{x}) \log \frac{f_\mathbf{x}(\mathbf{x})}{g_\mathbf{x}(\mathbf{x})} d\mathbf{x}. \tag{A5}
\]

The Kullback–Leibler divergence is not symmetric and therefore it does not satisfy the distance axioms. However, because \( D_{\text{KL}}[f_\mathbf{x}(\mathbf{x}), g_\mathbf{x}(\mathbf{x})] \geq 0 \) with equality if and only if \( f_\mathbf{x}(\mathbf{x}) = g_\mathbf{x}(\mathbf{x}) \), it can be used as a measure of quantifying the closeness of two distributions. Comparison between (A4) and (A5) reveals that the mutual information between \( \hat{s}_1, \ldots, \hat{s}_k \) is identical to the KL divergence between \( f_\mathbf{x}(\mathbf{s}) \) and its independence version \( \Pi_{i=1}^k f_\mathbf{x}(\hat{s}_i) \); that is, \( \phi_{\text{MI}}[\mathbf{s}] = D_{\text{KL}}[f_\mathbf{x}(\mathbf{s}), \Pi_{i=1}^k f_\mathbf{x}(\hat{s}_i)] \).

Among all distributions with fixed covariance structure, the normal distribution maximizes the differential entropy (Cover and Thomas 1991). Source separation may be achieved by exploiting this property of the normal distribution. In light of (A4), minimizing the mutual information between \( \hat{s}_1, \ldots, \hat{s}_k \) is equivalent to minimizing the marginal entropies \( \sum_{i=1}^k H[\hat{s}_i] \), which in turn amounts to maximizing their departure from normality. The deviation from normality of \( \hat{s} \) may be quantified conveniently in terms of the negentropy measure (Lee 1998). Let \( f_{\mathbf{x}}(\mathbf{x}) \) be a Gaussian pdf with entropy \( H[\mathbf{x}] \) having the same mean and covariance matrix as \( f_\mathbf{x}(\mathbf{x}) \). Then, negentropy can be defined as

\[
J(\mathbf{x}) = D_{\text{KL}}[f_\mathbf{x}(\mathbf{x}), f_{\mathbf{x}}(\mathbf{x})] = H[\mathbf{x}] - H[\mathbf{x}] \geq 0. \tag{A6}
\]

For sphered data, negentropy is related to mutual information by

\[
I[\mathbf{x}] = J(\mathbf{x}) - \sum_{i=1}^k J(x_i). \tag{A7}
\]

The use of entropies or negentropy to approximate independence requires estimates of the densities involved. This is computationally rather complicated and/
or error prone. It is common in practice to approximate these contrasts using higher-order cumulants (Cardoso 1999). Approximations of the contrasts through higher-order statistics (or cumulants) can be obtained using polynomial representations of the pdfs in an orthonormal series expansion (Stuart and Ord 1994) [see also Lee (1998) for an approximation of $\phi$[8]]:

APPENDIX B

Gradient of the ICA Rotation Objective Function

Given an $n \times k$ data matrix $Y = (y_{ij})$ (e.g., PCs) in which the columns and rows represent the variables and observations (or time index), respectively, let $V$ be a $k \times k$ rotation matrix and

$$ G = YV, \quad (B1) $$

the rotated data matrix. We also let

$$ L = G \otimes G = G^2, \quad (B2) $$

the component-wise square of $G$ and $R = L^T H L$, the covariance matrix of $L = (L_{ij})$, with $\otimes$ being the component-wise multiplication operator, and $H = (h_{ij})$ the centering operator [see Eq. (20)]. The aim is to compute the gradient of the objective function

$$ f(V) = \text{trace}[R^T (R \otimes N)] = ||R||_F - ||\text{diag}(R)||_F, \quad (B3) $$

where $||\cdot||$ stands for the Frobenius norm [see Eq. (21)], $N = 1_{k \times k} - I_k$ is the $k \times k$ matrix containing zeros in the main diagonal and ones elsewhere, and $\text{trace}(\cdot)$ stands for the trace operator. To compute the gradient of (A3) we use the chain rule.

First we consider the function

$$ f_0(X) = \text{trace}[X^T (X \otimes N)]. \quad (B4) $$

We let the elements of $R$ and $N$ be $(r_{ij})$ and $(n_{ij})$, respectively. We first compute the elements of the operator $f_0(R)$. If $S(R) = R^T (R \otimes N) = (S_{ij})$, then

$$ S_{ij} = \Sigma_k r_{ik} r_{kj} n_{ki}, \quad (B5) $$

and

$$ f_0(R) = \sum_{ik} r_{ik}^2 n_{ki}. \quad (B6) $$

Now $\partial / \partial r_{ij} \text{trace}[S(R)] = \sum_{jk} r_{jk}^2 \delta_{kj} n_{ki} = 2 r_{ij} n_{ij}$ and

$$ \partial \text{trace}[S(R)] / \partial R = 2R \otimes N. \quad (B7) $$

Next we take the second part of the chain, namely, $R = R(L) = L^T H L$, and we let $q(L) = f_0[R(L)] = \text{trace}[(L^T H L) (L^T H L) \otimes N]$. Because $r_{ij} = L_{ik} h_{km} L_{mj}$, and $d q / d L_{ij} = \sum_{km} h_{km} (L_{mj} \delta_{ik} + L_{ma} \delta_{jk}) = [H L_{ia} \delta_{ja} + H L_{ja} \delta_{ji}]$, use the chain rule:

$$ \frac{\partial q}{\partial L_{ij}} = \sum_{ab} \frac{\partial f_0}{\partial r_{ab}} [R(L)] \frac{\partial r_{ab}}{\partial L_{ij}}, $$

where the first part in the rhs of the equality has been calculated in the first step above; that is, $\partial f_0 / \partial L_{ab} |_{R(L)} = 2[(L^T H L) \otimes N]_{ab}$. Therefore, after few lines of algebra, we get

$$ \frac{\partial q}{\partial L_{ij}} = 4 \sum_{\beta} \text{trace}[H L_{i\beta} (L^T H L) \otimes N]_{\beta j}; \quad (B8) $$

In the last step, we now have $\mathcal{L}(G) = G \otimes G = G^2$, and we let $q(G) = q(G^2)$, which yields

$$ \frac{\partial g}{\partial G} (G) = 2G \otimes \nabla_q[G^2]; \quad (B9) $$

Finally, keeping in mind that $f(V) = g(YV)$ and $\partial f(V) / \partial (V) = Y \nabla g(YV)$, one obtains

$$ \nabla f(V) = 8Y^T (YV \otimes \{H(YV)^2\} [Y(V)^2 H(YV)^2 \otimes N]). \quad (B10) $$

Note that so far in the above expression, the property of orthogonality is not yet taken into account. Let $O(k)$ be the manifold of all $k \times k$ orthogonal matrices; then,

$$ O(k) : = \{ V \in \mathbb{R}^{k \times k} | V V^T = V^T V = I_k \}. \quad (B11) $$

Given a current value of $V$, the gradient projection algorithm computes the gradient $\mathcal{F} (\cdot) = (1/2) f (\cdot)$ [see Eq. (21)], at $V$ first; that is, $\nabla V \mathcal{F} = 4Y^T \{G \otimes [H (G \otimes G)(N \otimes N)]\}$. The final expression of the gradient $\pi_{\tau}(V)$ of the rotation criterion at the $k \times k$ orthogonal matrix $V$ is obtained by projecting $\nabla V \mathcal{F}$ onto the linear manifold $T_V O(k)$ tangent at $V$ to $O(k)$. Now, it can be shown (e.g., Tendalov 2006) that

$$ \pi_{\tau}(V) = \frac{1}{2} [\nabla V \mathcal{F} - (\nabla V \mathcal{F})^T V]. \quad (B12) $$

This is a familiar general procedure to compute gradients of orthogonal rotation criteria (see, e.g., Jennrich 2001). The algorithm proceeds iteratively, is strictly descending, and converges from any starting point to a stationary point. At a stationary point $V$ of $\mathcal{F}$ restricted
REFERENCES


