A fully matricial approach to SVD, EOF and CCA

Jesús Fernández, Jon Sáenz, Juan Zubillaga

29th November 2002

1 Introduction

There is a large number of references about the multivariate statistical techniques to be described here [2, 3, 5, 6, 7]. Once said this, there is a question that should be answered at this point: Why another paper talking about something that has been explained so many times? Well, normally, the three techniques described here are not described in the same paper, and the people is prone to make use of the same letters to describe different things so when one tries to learn all of them has to keep one notation and make up another two of them. Here, the effort of making up new letters to refer to all the *objects* that appear in SVD, PCA and CCA has already been made. Of course, the notation doesn't match with any other publication. These lines are for the people interested in calculate numerically the vectors and time series related to this techniques. The final equations can be found inside with a new perspective: matrices. The calculation of these analysis is normally carried out with an indicial notation, but all of them are linear so can be expressed as matricial operations. With this approach the summation symbols disappear, the equations are more compactly written and the demonstrations are shorter. Besides, a lot of modern programming languages are able to carry out operations with matrices without taking care of each element separately. Another reason for writing these lines is to show how all you need to work with these techniques is an algorithm to perform the singular value decomposition of a matrix (and, luckily, this is present in any linear algebra package). The approach of the singular value decomposition is more stable [4, 5] than the usual diagonalization, that can lead to instability problems because of the non–full–rank matrices that usually appear.

2 Notation

Throughout the paper an indicial notation and a matricial one will be used. Matrices are presented as uppercase letters. The same matrix in indicial notation is represented as the corresponding lowercase letter.

When the time is involved as one of the dimensions of a matrix it is always the first index (rows). This is the case, for example, in a physical field defined in the space–time which will be presented as $\Phi_x(t_i, x_j)$ in indicial notation and just X in matricial form with each row containing the values of the field for each time step. This field is an exception to the rule of writing the indicial notation with the lowercased matrix name (this is done to avoid confusion with the x spatial coordinate). If another field is required, it will be referred to as $\Phi_y(t_i, y_k)$ or Y. In both cases, the fields are assumed to have zero mean without loss of generality. This means that they are not exactly fields but anomalies. The temporal mean at each site or grid point (each column of the matrix) is supposed to have already been removed.

The total number of time steps will be referred to as T, ie: the i in the previous t_i runs from 1 to T. Similarly for the space coordinates: $x_j, j = 1 \ldots N_x$ and $y_k, k = 1 \ldots N_y$

Other matrices are not representing space–time fields but one (or two) of the dimensions represent coordinates in a space different from the real one. In this case the non–space–time dimension is supposed to be the last one (ie: the columns). This will be the case with the EOFs $e_p(x_j)$, which matrix E has real–space (x_j) rows and EOF–space columns.

Not all equations are presented in both notations. As the text goes ahead the indicial notation is reduced to make the demonstrations shorter.

3 Singular Value Decomposition

The singular value decomposition has a twofold meaning. First, it's a matricial operation by which any matrix M can be decomposed into the product:

$$M = L \Sigma R^{\mathrm{T}} \tag{1}$$

where Σ is a diagonal matrix and L and R are matrices with orthonormal vectors as columns [5]. This matrices are not strictly orthogonal because they are not square matrices (this kind of matrices are sometimes referred to semiorthogonal). Anyway, the orthonormality of the vectors imply $L^{\mathrm{T}} L = \mathbb{I}$ and $R^{\mathrm{T}} R = \mathbb{I}$. And second, it's a multivariate statistical decomposition technique by which two fields can be expanded in series in such a way that each term of the summation has maximum covariance with the correspondent one of the other field.

$$\begin{aligned} \Phi_x(t_i, x_j) &= \sum_p j_p(t_i) l_p(x_j) \\ \Phi_y(t_i, y_k) &= \sum_p k_p(t_i) r_p(y_k) \\ X &= J L^{\mathrm{T}} \\ Y &= K R^{\mathrm{T}} \end{aligned} \tag{2}$$

The coincidence in the nomenclature is easily understood because it can be proved [2] that the vectors $l_p(x_j)$ and $r_p(y_k)$ with maximum covariance are the singular vectors (the columns of the matrices L and R) of the cross–covariance matrix of the fields:

$$
\begin{aligned}
c_{xy}(x_j, y_k) &= \frac{1}{T}\sum_i \Phi_x(t_i, x_j)\Phi_y(t_i, y_k) \\
C_{xy} &= \frac{1}{T}X^{\mathrm{T}}Y
\end{aligned} \tag{3}
$$

And the singular coefficients $j_p(t_i)$ and $k_p(t_i)$ are obtained from the equations (2) just taking into account the orthonormality of the singular vectors:

$$
\begin{aligned}
j_p(t_i) &= \sum_j \Phi_x(t_i, x_j) l_p(x_j) \\
k_p(t_i) &= \sum_k \Phi_y(t_i, y_k) r_p(y_k) \\
J &= XL \\
K &= YR
\end{aligned} \tag{4}
$$

The (maximized) covariance of each term in the equations (2) is proved to be the corresponding singular value:

$$
\begin{aligned}
c_{jk}(j_p, k_q) &= \frac{1}{T}\sum_i j_p(t_i) k_q(t_i) \\
&= \frac{1}{T}\sum_{i,j,k} \Phi_x(t_i, x_j) l_p(x_j) \Phi_y(t_i, y_k) r_q(y_k) \\
&= \sum_{j,k} l_p(x_j) c_{xy}(x_j, y_k) r_q(y_k) \\
&= \sum_{j,k,p'} l_p(x_j) l_{p'}(x_j) \sigma_{p'p'} r_{p'}(y_k) r_q(y_k) \\
&= \sum_{p'} \delta_{pp'} \sigma_{p'p'} \delta_{p'q} \\
&= \sigma_{pq}\delta_{pq} \\
C_{jk} &= \frac{1}{T}J^{\mathrm{T}}K = \frac{1}{T}L^{\mathrm{T}}X^{\mathrm{T}}YR = L^{\mathrm{T}}C_{xy}R = L^{\mathrm{T}}L\Sigma R^{\mathrm{T}}R = \Sigma
\end{aligned} \tag{5}
$$

This is a very good example of how the matricial approach simplifies the calculus and notation.

4 Empirical Orthogonal Functions

The calculus of empirical orthogonal functions (EOFs) is also called Principal Component Analysis (PCA). It's another multivariate technique to decompose a field. In this case it is applied to just a single field and the result is a series with the terms explaining the maximum possible variance of the original field.

The temporal coefficients of this decomposition $\alpha_p(t_i)$ are usually referred to as principal components (PCs) and the spatial patterns $e_p(x_j)$ empirical orthogonal functions (EOFs). These EOFs represent the linear combination of the original variables which explain the maximum amount of variance of the original field in decreasing order. The field $\Phi_x(t_i, x_j)$ can be recovered as:

$$\begin{aligned} \Phi_x(t_i, x_j) &= \sum_p \alpha_p(t_i) e_p(x_j) \\ X &= \mathcal{A} E^{\mathrm{T}} \end{aligned} \tag{6}$$

It can be proved that the maximum variance is explained by choosing the EOFs as the eigenvectors of the covariance matrix $C_{xx} = \frac{1}{T} X^{\mathrm{T}} X$:

$$\begin{aligned} \sum_k c_{xx}(x_j, x_k) e_p(x_k) &= \lambda_p e_p(x_j) \\ C_{xx} E &= E \Lambda \end{aligned} \tag{7}$$

As the covariance matrix is definite positive and symmetric, the eigenvectors can be chosen orthonormal. In such a case the eigenvalues λ_p are the variances of the PCs which are obtained by projecting the field onto the EOFs ($\mathcal{A} = XE$, just taking into account the orthonormality of E)

$$C_{\alpha\alpha} = \frac{1}{T} \mathcal{A}^{\mathrm{T}} \mathcal{A} = \frac{1}{T} E^{\mathrm{T}} X^{\mathrm{T}} X E = E^{\mathrm{T}} C_{xx} E = E^{\mathrm{T}} E \Lambda = \Lambda \tag{8}$$

There is no temporal covariance between different PCs, that is, the PCs are uncorrelated. Each one explains a disjoint part of the total variance. The total variance of the field can be computed adding up all the eigenvalues λ_p.

The selection of orthonormalized EOFs is just one of the possibilities when scaling the EOFs. Other possibility is to standardize the PCs dividing each of them by their standard deviation, that is, the square root of the eigenvalues. This way, it can be defined a new set of PCs ($\alpha'_p(t_i)$) and EOFs ($e'_p(x_j)$) which also match the maximum explained variance rule:

$$\begin{aligned} \alpha'_p(t_i) &= \frac{\alpha_p(t_i)}{\sqrt{\lambda_p}} \\ e'_p(x_j) &= \sqrt{\lambda_p} e_p(x_j) \\ \mathcal{A}' &= \mathcal{A} \Lambda^{-1/2} \\ E' &= E \Lambda^{1/2} \end{aligned} \tag{9}$$

This way the PCs have no physical units and they are in the EOFs. The new EOFs are no longer orthonormal, but only orthogonal. Obviously, this transformation can only be done if none of the eigenvalues λ_p (or at least none of the ones you are retaining) is zero. There is also the possibility of using standardized PCs and orthonormal EOFs, but the series used to reconstruct the original field must include in this case the standard deviations (the square roots of the eigenvalues):

$$\begin{aligned}\Phi_x(t_i,x_j) &= \sum_p \alpha_p(t_i)e_p(x_j) = \sum_p \alpha'_p(t_i)e'_p(x_j) = \\ &= \sum_p \sqrt{\lambda_p}\alpha'_p(t_i)e_p(x_j) \\ X &= \mathcal{A}E^{\mathrm{T}} = \mathcal{A}'E'^{\mathrm{T}} = \\ &= \mathcal{A}'\Lambda^{1/2}E^{\mathrm{T}}\end{aligned} \tag{10}$$

4.1 SVD approach to the EOF calculus.

The covariance matrix of the field is unlikely to be of full rank (this would only be the case if the number of temporal samples is higher than the number of space channels and the spatial grid points are completely independent), therefore, there will be several eigenvalues that could be zero. Some routines crash due to the degeneracy of the eigenvalue that do not completely define the associated eigenvector. A numerically more stable [4, 5] way of computing the EOFs is through the previously mentioned SVD matricial decomposition. Consider the SVD decomposition of the anomaly field:

$$X \overset{\mathrm{SVD}}{=} L\Sigma R^{\mathrm{T}} \tag{11}$$

This has nothing to do with the SVD multivariate statistical technique described in section (3). In that case the decomposed matrix was the cross–covariance matrix of two fields. In this case we are considering just one field and the decomposed matrix is directly the anomaly field.

Now, the covariance matrix can be written

$$C_{xx} = \frac{1}{T}X^{\mathrm{T}}X = \frac{1}{T}R\Sigma L^{\mathrm{T}}L\Sigma R^{\mathrm{T}} = \frac{1}{T}R\Sigma^2 R^{\mathrm{T}} \tag{12}$$

And also through the equation (7) we can write

$$C_{xx} = E\Lambda E^{\mathrm{T}} \tag{13}$$

Comparing these two equations we have

$$\begin{aligned} E &= R \\ \Lambda &= \frac{1}{T}\Sigma^2 \end{aligned} \tag{14}$$

And the PCs are also obtained from SVD components:

$$\mathcal{A} = XE = (L\Sigma R^{\mathrm{T}})R = L\Sigma \tag{15}$$

Standardized PCs can also be obtained in this way, just using L, R and Σ from the same SVD decomposition of equation (11).

$$\begin{aligned} E' &= \frac{1}{\sqrt{T}} R\Sigma \\ \mathcal{A}' &= \sqrt{T} L \end{aligned} \tag{16}$$

5 Canonical Correlation Analysis.

The canonical correlation analysis (CCA) is another technique to decompose two fields. In this case it is done in such a way that the correlation between the temporal expansion coefficients ($a_p(t_i)$ and $b_p(t_i)$) is maximum. This time the spatial patterns ($p_p(x_j)$ and $q_p(y_k)$) are not even orthogonal, so the matricial computations must be done with care.

$$\begin{aligned} \Phi_x(t_i, x_j) &= \sum_p a_p(t_i) p_p(x_j) \\ \Phi_y(t_i, y_k) &= \sum_p b_p(t_i) q_p(y_k) \\ X &= AP^{\mathrm{T}} \\ Y &= BQ^{\mathrm{T}} \end{aligned} \tag{17}$$

The concept of adjoint pattern is necessary to deal with this non–orthogonal patterns. The adjoint pattern is defined by the following relationship:

$$\begin{aligned} \sum_j p_p(x_j) {p_q}^{\mathrm{A}}(x_j) &= \delta_{pq} \\ P^{\mathrm{T}} P^{\mathrm{A}} &= \mathbb{I} \end{aligned} \tag{18}$$

With this adjoint patterns the expressions for the temporal expansion coefficients are immediate

$$\begin{aligned} a_p(t_i) &= \sum_j \Phi_x(t_i, x_j) {p_p}^{\mathrm{A}}(x_j) \\ b_p(t_i) &= \sum_j \Phi_y(t_i, y_j) {q_p}^{\mathrm{A}}(y_j) \\ A &= XP^{\mathrm{A}} \\ B &= YQ^{\mathrm{A}} \end{aligned} \tag{19}$$

It can be demonstrated [8] that the maximum correlation constraint is filled if the canonical patterns are taken as the eigenvectors of the following matrices M_x and M_y

$$
\begin{aligned}
M_x &= C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \\
M_y &= C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}
\end{aligned} \tag{20}
$$

that is,

$$
\begin{aligned}
\sum_k m_x(x_j, x_k) p_p{}^{\mathrm{A}}(x_k) &= \mathfrak{r}_p^2 p_p{}^{\mathrm{A}}(x_j) \\
\sum_k m_y(y_j, y_k) q_p{}^{\mathrm{A}}(y_k) &= \mathfrak{r}_p^2 q_p{}^{\mathrm{A}}(y_j) \\
M_x P &= P \mathfrak{R}^2 \\
M_y Q &= Q \mathfrak{R}^2
\end{aligned} \tag{21}
$$

By construction the eigenvalues $\mathfrak{r}_p^2$ are equal for both matrices and are the squares of the canonical correlations, that is, the correlation between the canonical coefficients (this will be more easily demonstrated later). Once the patterns have been obtained the canonical coefficients are calculated through equation (19). The problem of the indetermination of the coefficients by a constant factor arises again in this technique. This time it is no possible to take orthonormal spatial patterns so the most logical choice is to take standardized coefficients and, consequently, spatial patterns with the physical units of the field.

5.1 The Barnett–Preisendorfer approach to CCA.

Barnett and Preisendorfer [1] proposed to filter the fields with an EOF decomposition prior to enter the CCA and carry out the calculation in EOF coordinates. There are several reasons for doing so. The noise of the data is filtered out by truncating the EOF series. In the EOF–space the coordinates of the field are the PCs and when using standardized PCs the auto–covariance matrices that appear in the definition of the M_x and M_y matrices are equal to the identity matrix largely simplifying the calculus. The resulting M_x and M_y matrices are symmetric and, consequently, the canonical patterns in the EOF–space (P_{EOF} and Q_{EOF}) are orthonormal. Apart from the Barnett–Preisendorfer suggestion we are also going to make use of the SVD matricial decomposition to solve the problem of finding the CCA patterns and coefficients.

Let's consider the EOF decomposition of the two anomaly fields (with standardized PCs):

$$
\begin{aligned}
\Phi_x(t_i, x_j) &= \sum_p \alpha'_p(t_i) e'_p(x_j) \\
\Phi_y(t_i, y_k) &= \sum_p \beta'_p(t_i) f'_p(y_k) \\
X &= \mathcal{A}' E'^{\mathrm{T}} \\
Y &= \mathcal{B}' F'^{\mathrm{T}}
\end{aligned} \tag{22}
$$

If we write the equation (21), that defines the canonical patterns, in EOF–coordinates we get

$$\begin{aligned} C'_{\alpha\beta}C'_{\beta\alpha}P_{\text{EOF}} &= P_{\text{EOF}}\mathfrak{R}^2 \\ C'_{\beta\alpha}C'_{\alpha\beta}Q_{\text{EOF}} &= Q_{\text{EOF}}\mathfrak{R}^2 \end{aligned} \tag{23}$$

after taking into account that for the standardized PCs the covariance matrices $C'_{\alpha\alpha}$ and $C'_{\beta\beta}$ are identity matrices. Now, performing the SVD decomposition of the matrix $C'_{\alpha\beta}$

$$C'_{\alpha\beta} \overset{\text{SVD}}{=} L\Sigma R^{\text{T}} \tag{24}$$

we also get:

$$\begin{aligned} C'_{\alpha\beta}C'_{\beta\alpha} &= C'_{\alpha\beta}{C'_{\alpha\beta}}^{\text{T}} = L\Sigma R^{\text{T}}R\Sigma L^{\text{T}} = L\Sigma^2L^{\text{T}} \\ C'_{\beta\alpha}C'_{\alpha\beta} &= {C'_{\alpha\beta}}^{\text{T}}C'_{\alpha\beta} = R\Sigma L^{\text{T}}L\Sigma R^{\text{T}} = R\Sigma^2R^{\text{T}} \end{aligned} \tag{25}$$

or, stated in other way,

$$\begin{aligned} C'_{\alpha\beta}C'_{\beta\alpha}L &= L\Sigma^2 \\ C'_{\beta\alpha}C'_{\alpha\beta}R &= R\Sigma^2 \end{aligned} \tag{26}$$

which compared to equations (23) yields:

$$\begin{aligned} P_{\text{EOF}} &= L \\ Q_{\text{EOF}} &= R \end{aligned} \tag{27}$$

and remembering that both Σ and $\mathfrak{R}$ are definite positive matrices we also have

$$\mathfrak{R}^2 = \Sigma^2 \Longrightarrow \mathfrak{R} = \Sigma \tag{28}$$

Now, to go back to the real space we only need to use the EOFs matrix to invert the transformation:

$$\begin{aligned} P &= E'P_{\text{EOF}} = E'L \\ Q &= F'Q_{\text{EOF}} = F'R \end{aligned} \tag{29}$$

As the L and R matrices come from the SVD decomposition of a non–dimensional matrix (because it is the covariance matrix of two non–dimensional PCs) they also have no physical dimensions. The E' and F' EOF matrices have the dimensions of the original field as discussed in the end of section 4. Therefore, the canonical patterns obtained through (29) also have the units of the field. The adjoint patterns must have the inverse units so that the equation (18) holds. This equation is trivially proved to be held by the following adjoint patterns:

$$\begin{aligned} P^{\text{A}} &= E'\Lambda_x^{-1}L \\ Q^{\text{A}} &= F'\Lambda_y^{-1}R \end{aligned} \tag{30}$$

And the canonical coefficients are

$$\begin{aligned} A &= XP^{\wedge} = XE'\Lambda_x^{-1}L = XE\Lambda_x^{-1/2}L = \mathcal{A}\Lambda_x^{-1/2}L = \mathcal{A}'L \qquad (31) \\ B &= YQ^{\wedge} = YF'\Lambda_y^{-1}R = YF\Lambda_y^{-1/2}R = \mathcal{B}\Lambda_y^{-1/2}R = \mathcal{B}'R \end{aligned}$$

These are easily proved to have unit variance, for instance the auto–covariance matrix for A is

$$Cov(A, A) = \frac{1}{T}A^{\mathrm{T}}A = \frac{1}{T}L^{\mathrm{T}}\mathcal{A}'^{\mathrm{T}}\mathcal{A}'L = L^{\mathrm{T}}L = \mathbb{I} \qquad (32)$$

and its diagonal is made up with the variance of each PC. As the canonical coefficients are standardized their correlation is just the cross–covariance matrix

$$\begin{aligned} Corr(A, B) &= Cov(A, B) = \frac{1}{T}A^{\mathrm{T}}B = \frac{1}{T}L^{\mathrm{T}}\mathcal{A}'^{\mathrm{T}}\mathcal{B}'R \qquad (33) \\ &= L^{\mathrm{T}}C'_{\alpha\beta}R = L^{\mathrm{T}}L\Sigma R^{\mathrm{T}}R = \Sigma \end{aligned}$$

This result confirms our previous assertion that the diagonal matrix of the EOF–space cross–covariance matrix SVD decomposition is the canonical correlation matrix.

6 Concluding remarks

The three techniques described here show a lot of similarities. This is not surprising because the underlying idea is common to all of them: maximize an statistic of one or more fields when their series expansion is truncated to the leading terms. It has already been pointed out that the statistic is the cross–covariance for SVD, the auto–covariance for PCA and the cross–correlation for CCA. There is still another technique related to PCA that maximizes the auto–correlation. This PCA is achieved by just standardizing the field prior to carry out the calculations showed in the text.

These statistics are linearly calculated out of the field values. This fact along with the linearity of a series expansion makes possible a fully matricial description of this techniques, as has already been demonstrated.

The most common approach to carry out these calculations is the diagonalization of the involved matrices, obtaining the patterns and the expansion coefficients from the eigenvalues and eigenvectors. This approach is not the most stable and another one has been introduced: the singular value decomposition. This is kind of a generalization of the eigentechniques for the case of non–square matrices.

References

[1] T. P. Barnett and R. Preisendorfer. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, 115:1825–1850, 1987.

[2] C. S. Bretherton, C. Smith, and J. M. Wallace. An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, 5:541–560, 1992.

[3] S. Cherry. Singular value decomposition analysis and canonical correlation analysis. *J. Climate*, 9:2003–2009, 9 1996.

[4] William J. Emery and Richard E. Thomson. *Data Analysis Methods in Physical Oceanography.* Pergamon, 1998.

[5] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, London, 1996.

[6] M. Newman and P.D. Sardeshmukh. A caveat concerning singular value decomposition. *J. Climate*, 8:352–360, 1995.

[7] H. von Storch and A. Navarra, editors. *Analysis of Climate Variability. Applications of Statistical Techniques.* Springer, Berlin, 1995.

[8] D. S. Wilks. *Statistical methods in the atmospheric sciences.* Academic Press, 1995.