Some Comments on Singular Value Decomposition Analysis

STEVE CHERRY

Department of Mathematical Sciences, Montana State University, Bozeman, Montana

(Manuscript received 26 January 1996, in final form 7 June 1996)

ABSTRACT

The singular value decomposition analysis (SVD) method is discussed in the context of the simultaneous orthogonal rotation of two matrices. It is demonstrated that the singular vectors are rotated EOFs and the SVD expansion coefficients are rotated sets of principal component expansion coefficients. This way of thinking about SVD aids in the interpretation of results and provides guidance as to when and how to use SVD.

1. Introduction

The singular value decomposition (SVD) method of detecting temporally synchronous spatial patterns has been popularized by Bretherton et al. (1992) and Wallace et al. (1992). The goal of SVD is to find linear combinations of two sets of variables such that the linear combinations have the maximum possible covariance. The maximization is carried out under orthogonality constraints on the coefficients of the linear combinations. The method is based on a singular value decomposition of the matrix whose elements are covariances between observations made at different grid points in two geophysical fields. Examples of its use in the atmospheric sciences can be found in Prohaska (1976), Lanzante (1984), Wallace et al. (1992), Hsu (1994), and Lau and Nath (1994). It is apparent that SVD has become a popular method of data analysis in the atmospheric sciences, in part because of its simplicity.

In a recent paper, Newman and Sardeshmukh (1995) discussed some of the limitations of SVD and some of the hidden assumptions that investigators make when they use the method. They argued that SVD will only be capable of detecting coupled patterns under very special circumstances. Cherry (1996) has also raised questions about the usefulness of SVD.

Although relatively new to atmospheric scientists, SVD has been practiced for many years in the social sciences (Tucker 1958; Cliff 1966; Van de Geer 1984). In an unpublished paper, K. E. Muller (1982, personal communication) referred to the method as canonical covariance analysis and argued that it could be characterized as one of finding orthogonal transformations of principal components scores (i.e., expansion coefficients) from covariance matrices.

The goal of this paper is to show that SVD is essentially a solution to a particular matrix matching problem. K. E. Muller's (1982, personal communication) comment connecting principal components analysis (PCA) and SVD will be explored. This has implications for interpreting the results of an SVD analysis. This paper is organized as follows. Section 2 contains a brief description of SVD. In section 3, the problem of finding two orthogonal matrices that rotate two data matrices to congruence (in a least squares sense) will be described, and in section 4, the connection to PCA will be discussed. Some conclusions will be drawn in section 5.

2. Singular value decomposition analysis

Let $\mathbf{X}$ and $\mathbf{Y}$ be $n \times p$ and $n \times q$ data matrices. It will be further assumed that the means of the columns of $\mathbf{X}$ and $\mathbf{Y}$ are all equal to 0. Let $\mathbf{C}_{xy} = (1/n)\mathbf{X}'\mathbf{Y}$ be the $p \times q$ matrix, whose elements are the covariances between the time series in the two fields.

The goal of SVD is to find linear combinations of the data $\mathbf{X}\mathbf{a}_i$ and $\mathbf{Y}\mathbf{b}_i$ [$i = 1, \ldots, r$; $r = \min(p, q)$], with the maximum covariance, subject to the $p \times 1$ vectors $\mathbf{a}_i$ and $q \times 1$ vectors $\mathbf{b}_i$, satisfying the orthogonality constraints

$$\mathbf{a}_i'\mathbf{a}_j = \mathbf{b}_i'\mathbf{b}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

The solution is found by taking the singular value decomposition of $\mathbf{C}_{xy}$, denoted here by $\mathbf{C}_{xy} = \mathbf{ADB}'$; $\mathbf{A}$ is a $p \times r$ semiorthogonal matrix, $\mathbf{D}$ is an $r \times r$ diagonal matrix, and $\mathbf{B}$ is a $q \times r$ semiorthogonal matrix. The ith columns of $\mathbf{A}$ and $\mathbf{B}$ contain the left and right weight (singular) vectors $\mathbf{a}_i$ and $\mathbf{b}_i$, and the ith element of $\mathbf{D}$ (the ith singular value) is the covariance of $\mathbf{X}\mathbf{a}_i$ and $\mathbf{Y}\mathbf{b}_i$. The vectors $\mathbf{X}\mathbf{a}_i$ and $\mathbf{Y}\mathbf{b}_i$ are referred to as the ith pair of SVD expansion coefficients.

Corresponding author address: Dr. Steve Cherry, Dept. of Mathematical Sciences, Montana State University, Bozeman, MT 59717-0240.
E-mail: imsgsche@math.montana.edu

The two sets of expansion coefficients (which are time series) will tend to be correlated with one another because

$$\mathrm{cov}(\mathbf{X}\mathbf{a}_i, \mathbf{Y}\mathbf{b}_i) = \mathrm{cor}(\mathbf{X}\mathbf{a}_i, \mathbf{Y}\mathbf{b}_i)[\mathrm{var}(\mathbf{Y}\mathbf{a}_i)\mathrm{var}(\mathbf{Y}\mathbf{b}_i)]^{1/2}. \quad (1)$$

Thus, finding $\mathbf{a}_i$ and $\mathbf{b}_i$ to maximize the covariance of the expansion coefficients will tend to produce pairs of expansion coefficients that are correlated. The hope then is that the pairs of patterns seen in the contour maps are geophysically meaningful and synchronous in time. It should also be kept in mind that there is always a mathematical solution to the SVD problem, and there is always a chance that a pair of coupled patterns that seem geophysically relevant is nothing more than a mathematical artifact (Cherry 1996).

3. Simultaneous orthogonal rotation to congruence

A common problem in psychometrics is to find matching patterns in two seemingly different sets of data. For example, a collection of individuals may be given two different batteries of tests. The investigator believes (or hopes) that the two batteries are measuring the same thing, but the two sets of data appear to be quite different. The goal is to rotate the two sets so that they are as similar as possible. Cliff (1966) presented a solution to this problem when the rotations were restricted to be orthogonal.

As above, let $\mathbf{X}$ and $\mathbf{Y}$ be two mean centered data matrices of sizes $n \times p$ and $n \times q$. It is desirable to find the directions in p space and q space so that when the variables in the two sets of data are projected onto these two axes, they are as similar as possible. The next step is to find a second set of directions, orthogonal to the first, with similar properties, and so on up to r such pairs of directions. Stated another way, we seek a $p \times r$ matrix $\mathbf{A}$ and a $q \times r$ matrix $\mathbf{B}$ such that the quantity

$$\|\mathbf{XA} - \mathbf{YB}\|^2 \quad (2)$$

is minimized subject to $\mathbf{A}$ and $\mathbf{B}$ being semiorthogonal. The notation $\|\ \|$ refers to the Frobenius norm of a matrix. If $\mathbf{G}$ is an $n \times p$ matrix with elements g_{ij}, then

$$\|\mathbf{G}\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} g_{ij}^2.$$

That is, the rotation matrices are attempting to match the ith observation (or row) of $\mathbf{X}$ with the ith observation (or row) of $\mathbf{Y}$, and this matching is done in such a way that the sum of the squared differences between the rotated points is as small as possible.

A solution to the problem is to let $\mathbf{A}$ and $\mathbf{B}$ be the matrices of left and right singular vectors from the singular value decomposition of $\mathbf{C}_{xy}$. One ends up with pairs of orthogonal directions that make the corresponding columns of the SVD expansion coefficients $\mathbf{XA}$ and $\mathbf{YB}$ as similar as possible (in terms of minimized least squares distances). Thus, the solution to the maximal covariance problem in SVD is the same as the solution to the simultaneous orthogonal rotation problem.

4. PCA and SVD

Bretherton et al. (1992) point out that if $\mathbf{X}$ and $\mathbf{Y}$ are the same, then SVD is equivalent to PCA. There is more of a relationship between the two methods than that, however. The nature of the relationship is explored in this section, using the results from the previous section.

Consider Eq. (1) above. Maximizing $[\mathrm{var}(\mathbf{X}\mathbf{a}_i)\,\mathrm{var}(\mathbf{Y}\mathbf{b}_i)]^{1/2}$ subject to the constraints imposed by SVD can be accomplished by performing separate PCAs on $\mathbf{X}$ and $\mathbf{Y}$. But, of course, there is no guarantee that $\mathrm{cor}(\mathbf{X}\mathbf{a}_i, \mathbf{Y}\mathbf{b}_i)$ will be high. In fact, if the two fields were not coupled, then one would not expect any significant correlation between the pairs of expansion coefficients from separate PCAs of the two fields. SVD then can be thought of as a simultaneous orthogonal rotation of PCA expansion coefficients to congruence. This argument is made more rigorously below.

Denote the singular value decomposition of $\mathbf{X}$ and $\mathbf{Y}$ by

$$\mathbf{X} = \mathbf{A}_x\mathbf{D}_x\mathbf{B}'_x$$
$$\mathbf{Y} = \mathbf{A}_y\mathbf{D}_y\mathbf{B}'_y.$$

Here, $\mathbf{B}_x$ and $\mathbf{B}_y$ are the eigenvectors of of the covariance matrices $\mathbf{C}_{xx} = (1/n)\mathbf{X}'\mathbf{X}$ and $\mathbf{C}_{yy} = (1/n)\mathbf{Y}'\mathbf{Y}$, respectively. The matrix of principal components expansion coefficients for $\mathbf{X}$ is

$$\mathbf{X}^* = \mathbf{X}\mathbf{B}_x$$
$$= \mathbf{A}_x\mathbf{D}_x.$$

Similarly, the matrix of principal components expansion coefficients for $\mathbf{Y}$ is $\mathbf{Y}^* = \mathbf{A}_y\mathbf{D}_y$.

Then, writing $\mathbf{C}_{xy}$ in terms of the singular value decompositions of $\mathbf{X}$ and $\mathbf{Y}$ gives

$$\mathbf{C}_{xy} = \frac{1}{n}\mathbf{X}'\mathbf{Y}$$
$$= \frac{1}{n}\mathbf{B}_x\mathbf{D}_x\mathbf{A}'_x\mathbf{A}_y\mathbf{D}_y\mathbf{B}'_y$$
$$= \frac{1}{n}\mathbf{B}_x\mathbf{X}^{*\prime}\mathbf{Y}^*\mathbf{B}'_y$$
$$= \mathbf{B}_x\mathbf{C}_{x^*y^*}\mathbf{B}'_y,$$

where $\mathbf{C}_{x^*y^*}$ is the cross-covariance matrix of the principal components expansion coefficients from $\mathbf{X}$ and $\mathbf{Y}$. Denoting the singular value decomposition of $\mathbf{C}_{x^*y^*}$ by

$$\mathbf{C}_{x^*y^*} = \mathbf{PDQ}'$$

and substituting into the above system of equations gives

$$\mathbf{C}_{xy} = \mathbf{B}_x\mathbf{PDQ}'\mathbf{B}'_y$$
$$= \mathbf{ADB}',$$

where $\mathbf{ADB}'$ is the singular value decomposition of $\mathbf{C}_{xy}$ because of the (almost) unique properties of the singular value decomposition operation. Thus, $\mathbf{A} = \mathbf{B}_x\mathbf{P}$ and $\mathbf{B} = \mathbf{B}_y\mathbf{Q}$. Also, note that $\mathbf{C}_{xy}$ and $\mathbf{C}_{x*y*}$ have the same singular values.

Now, consider the simultaneous orthogonal rotation problem presented above in terms of the principal component expansion coeffcients. That is, find orthogonal matrices $\mathbf{P}$ and $\mathbf{Q}$ to minimize

$$\|\mathbf{X}^*\mathbf{P} - \mathbf{Y}^*\mathbf{Q}\|^2.$$

The $\mathbf{P}$ and $\mathbf{Q}$ that minimize this under the orthogonality constraints come from the singular value decomposition of $\mathbf{C}_{x*y*}$. Furthermore, note that

$$\|\mathbf{X}^*\mathbf{P} - \mathbf{Y}^*\mathbf{Q}\|^2 = \|\mathbf{X}\mathbf{B}_x\mathbf{P} - \mathbf{Y}\mathbf{B}_y\mathbf{Q}\|^2$$
$$= \|\mathbf{XA} - \mathbf{YB}\|^2.$$

Thus, the SVD expansion coefficients $\mathbf{XA}$ and $\mathbf{YB}$ can be thought of as orthogonally rotated principal component expansion coefficients $\mathbf{X}^*$ and $\mathbf{Y}^*$, and the SVD weights ($\mathbf{A}$ and $\mathbf{B}$) are orthogonally rotated PCA weights. As mentioned above, K. E. Muller (1982, personal communication) actually characterized SVD as a method of orthogonally rotating principal component expansion coefficients.

5. Discussion and conclusions

The results presented above suggest a possible course of action with respect to the use of SVD. First, separate PCAs should be carried out on the two fields. If the two sets of expansion coefficients are strongly correlated and the patterns are geophysically relevant, then one has evidence of coupling. It can be argued that one has fairly strong evidence of coupling because the association has not been mathematically forced on the data.

If SVD is considered appropriate, the interpretation of the results may be aided by recognizing that SVD is a simultaneous orthogonal rotation of the PCA solutions. The interpretation should be based in part on a comparison of the results (e.g., correlation maps) from the PCA analysis with those from the SVD analysis. This is, in fact, what Wallace et. al. (1992) did, but they were contrasting the two approaches, not considering them as two parts of a single exploratory process.

K. E. Muller (1982, personal communication) presented SVD as an alternative to canonical correlation analysis (CCA). He argued that SVD has appeal when the two sets of variables have common units of measurement and when differences in variation (covariation) are informative. CCA tends to obscure these differences. A CCA solution may be affected by highly correlated but unimportant (in the sense of low variation and/or covariation) variables within a set. K. E. Muller (1982, personal communication) argued that the decision of when to use SVD versus CCA is similar to the decision of whether or not to carry out PCA on covariance or correlation matrices. The restriction that both sets of variables must have the same units of measurement can be relaxed to one in which the units of measurement are the same within each set, a situation common in the atmospheric sciences.

Another issue is the effect of the orthogonality constraints in the SVD solution on the interpretability of results. Orthogonality is the restriction that the data space is not allowed to change shape (Cliff 1966). For example, if

$$\|\mathbf{XA} - \mathbf{YB}\|^2 = 0,$$

then the only difference between $\mathbf{X}$ and $\mathbf{Y}$ is an arbitrarily imposed coordinate system. SVD might, then, be most useful for investigating similarities in datasets that are more directly comparable (e.g., a comparison of model data with observational data). The basic idea is that if orthogonal rotations are appropriate, then the two sets of variables are the *same* in some sense, an idea consistent with the results presented in Newman and Sardeshmukh (1995).

The mathematical results presented above may seem trivial, but the implications of those results are not. Finding linear combinations with maximum covariance may seem to be a geophysically relevant thing to do, but before carrying out such an analysis, investigators should ask themselves if simultaneously rotating two datasets to congruence makes geophysical sense. If it does not, then SVD is probably not appropriate.

Acknowledgments. This work was supported by the National Science Foundation through the Geophysical Statistics Project at the National Center for Atmospheric Research in Boulder, Colorado.

REFERENCES

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560.
Cherry, S., 1997: Singular value decomposition analysis and canonical correlation analysis. *J. Climate,* **9,** 2003–2009.
Cliff, N., 1966: Orthogonal rotation to congruence. *Psychometrika,* **31,** 33–42.
Hsu, H., 1994: Relationship between tropical heating and global circulation: Interannual variability. *J. Geophys. Res.,* **99,** 10 473–10 489.
Lanzante, J. R., 1984: A rotated eigenanalysis of the correlation between 700-mb heights and sea surface temperatures in the Pacific and Atlantic. *Mon. Wea. Rev.,* **112,** 2270–2280.
Lau, N., and M. J. Nath, 1994: A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere–ocean system. *J. Climate,* **7,** 1184–1207.
Newman, M., and P. D. Sardeshmukh, 1995: A caveat concerning singular value decomposition. *J. Climate,* **8,** 352–360.
Prohaska, J., 1976: A technique for analyzing the linear relationships between two meteorological fields. *Mon. Wea. Rev.,* **104,** 1345–1353.
Tucker, N. J., 1958: An inter-battery method of factor analysis. *Psychometrika,* **23,** 111–136.
van de Geer, J. P., 1984: Linear relations among *k* sets of variables. *Psychometrika,* **49,** 79–94.
Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate,* **5,** 561–576.