Title: Eigenvalues, Singular Value Decomposition

Name: Radu C. Cascaval

Affil./Addr.: Department of Mathematics, University of Colorado Colorado Springs,

1420 Austin Bluffs Parkway, Colorado Springs, CO 80919

radu@uccs.edu

Eigenvalues, Singular Value Decomposition

Synonyms

Eigenvalues = Proper Values, Auto Values,

Singular Value Decomposition = Principal Component Analysis

Glossary

Matrix: a rectangular tableau of numbers

Eigenvalues: a set of numbers (real or complex) intrinsic to a given matrix

Eigenvectors: a set of vectors associated to a matrix transformation

Singular Value Decomposition: A specific decomposition of any given matrix, useful in matrix analysis and its applications

Definition

Eigenvalues and Eigenvectors

Given a square $(n \times n)$ matrix A, a (complex) number λ is called an **eigenvalue** of A if there exists a nonzero n-dimensional column vector X such that

$$AX = \lambda X, \quad X \neq 0. \tag{1}$$

A vector X satisfying (1) is called an **eigenvector** of A corresponding to eigenvalue λ.

Singular Value Decomposition (SVD)

Given any rectangular matrix $(m \times n)$ matrix A, by singular value decomposition of the matrix A we mean a decomposition of the form $A = U\Sigma V^T$, where U and V are orthogonal matrices (representing rotations) and Σ is a diagonal matrix (representing a stretch).

Introduction

Matrix analysis is ubiquitous in mathematics and its applications. Due to the context of this encyclopedia, we restrict our discussion to real matrices only. Moreover, we restrict most of the considerations here to real eigenvalues ($\lambda \in \mathbb{R}$) and real eigenvectors ($X \in \mathbb{R}^n$). One of the common ways to represent eigenvalues and eigenvectors is to associate A with a (linear) transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ given by the left-multiplication $X \longmapsto AX$; then eigenvectors are precisely those vectors that get mapped parallel to themselves and eigenvalues are the factors by which these eigenvectors stretch under this transformation; see (1). In particular, $\lambda = 0$ is eigenvalue for a matrix A precisely when A has a nontrivial

kernel (or nullspace): $\ker(A) = \{X \,|\, AX = 0\} \neq \{0\}$. For each eigenvalue λ of A, the set of eigenvectors corresponding to λ form (when including the zero vector as well) the **eigenspace** of A corresponding to λ:

$$E_\lambda^A = \{X \,|\, AX = \lambda X\}.$$

The set of eigenvalues of the matrix A is referred to as the **spectrum** of A. Depending on the application, one may only consider the real spectrum (hence only those eigenvalues which are real). Some matrices do not have any real spectrum (e.g. the matrix $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ representing a rotation by 90° in $\mathbb{R}^2$,) while other matrices have the entire spectrum on the real axis, as is the case for (real) symmetric matrices (see below).

We may refer to eigenvectors satisfying (1) as right-eigenvectors, to distinguish them from left-eigenvectors, defined as (row) vectors Y satisfying

$$YA = \lambda Y, \quad Y \neq 0.$$

Left eigenvectors of A are nothing else but the (right) eigenvectors of the transpose matrix A^T. (The transpose B^T of a matrix B is defined as the matrix obtained by rewriting the rows of B as columns of the new B^T and viceversa.) While the eigenvalues of A and A^T are the same, the sets of left- and right- eigenvectors may be different in general.

Eigenvalue and Eigenvector Computation

We now describe how to find the eigenvalues of a given matrix. The eigenvalues of A turn out to be precisely the roots of the characteristic polynomial of the matrix A, $p_A(t) := \det(A - tI_n)$, where I_n is the identity $n \times n$ matrix:

$$\lambda \text{ is eigenvalue for A} \iff p_A(\lambda) = \det(A - \lambda I_n) = 0.$$

The Fundamental Theorem of Algebra guarantees that any polynomial with real coefficients, such as p_A, can be factored into linear factors

$$p_A(t) = (-1)^n(t-\lambda_1)^{m_1}\dots(t-\lambda_k)^{m_k}$$

where $\lambda_1,\dots\lambda_k$ are precisely the distinct (complex) eigenvalues of A. The positive integer m_j is called the algebraic multiplicity of the eigenvalue λ_j, $j = 1,\dots,k$. Non-real eigenvalues, if any, come in complex conjugate pairs. The other important information about each eigenvalue $\lambda = \lambda_j$ is its geometric multiplicity, which is defined as $\dim E^A_\lambda$, the dimension of the eigenspace E^A_λ, or the maximum number of linearly independent eigenvectors of A corresponding to λ. A well-known fact in linear algebra reads

$$\text{geometric multiplicity of } \lambda \leq \text{ algebraic multiplicity of } \lambda. \tag{2}$$

Matrices for which the above equality holds for each of its eigenvalues are called diagonalizable, since the matrix A can be represented as a diagonal matrix (see below).

Diagonalization of Symmetric Matrices

By definition, a $n \times n$ matrix $A = (a_{ij})$ is symmetric if $a_{ij} = a_{ji}$ for all indices $i, j = 1,\dots,n$, or, in short, if it equals its own transpose $A = A^T$. We describe here a fundamental property of symmetric matrices, which is that any symmetric matrix A has a decomposition of the form

$$A = SDS^T, \tag{3}$$

where S = orthogonal matrix, ($S^TS = I_n = SS^T$, or $S^{-1} = S^T$) and D = diagonal matrix. We say that A is diagonalizable and that S and D diagonalize the matrix A. A generalization of this decomposition for the case of a (possibly non-square) $m \times n$ matrix is precisely the SVD (see Section 3).

In the remaining of this section, we present the construction of the diagonalization procedure (3) for a symmetric $n \times n$ matrix. First, all eigenvalues of a (real) symmetric matrix are real. **Second, eigenvectors corresponding to distinct eigenvalues are orthogonal. [Two vectors X, Y in $\mathbb{R}^N$ are called orthogonal if $X^T Y = 0$.] A more substantial fact, fundamental in linear algebra, is that for symmetric matrices, the geometric multiplicity of each eigenvalue *equals* its algebraic multiplicity (equality in (2)), hence $\sum_{j=1}^{k} \dim E_{\lambda_j} = \sum_{j=1}^{k} m_j = n$. This translates into the fact that are sufficiently many (precisely n) linearly independent eigenvectors of A to form a basis of $\mathbb{R}^n$. [A basis in $\mathbb{R}^n$ is a set of n linearly independent vectors.] Finally, the n linearly independent eigenvectors of A can be chosen to be mutually orthogonal (using Gram-Schmidt orthogonalization process within each eigenspace, if necessary) and consequently, form an orthonormal basis of $\mathbb{R}^n$. [An orthonormal basis is a basis consisting of mutually orthogonal vectors which are also unit length]. A direct consequence of the above-mentioned facts is that the matrix S constructed by placing as its columns precisely the n eigenvectors described above, and the diagonal matrix D constructed by choosing as the diagonal entries precisely the eigenvalues λ_j, listed with their multiplicities m_j, fulfill the relation**

$$S^{-1}AS = D \text{ (diagonal matrix)} .$$

By design, S is an orthogonal matrix, i.e. satisfies $S^T S = I_n$, since the columns vectors for an orthonormal set in $\mathbb{R}^n$. Using the orthogonality of S, rewritten as $S^{-1} = S^T$, we then solve for A to obtain $A = SDS^{-1} = SDS^T$, which is the desired decomposition (3). Note that this decomposition is not unique, in general, since it depends on the choice of the eigenvectors used. The diagonalization of symmetric

matrices has a wide range of applications (in classical mechanics, dynamical systems etc). It is also the springboard towards a generalization for non-symmetric, even non-square matrices, which will be described in the next section.

Singular Value Decomposition (SVD)

Definition

We describe here a very important generalization of the results above. Precisely, any rectangular $(m \times n)$ matrix A with real coefficients, admits a decomposition of the form

$$A = U\Sigma V^T. \tag{4}$$

with U an orthogonal $(m \times m)$ matrix, V an orthogonal $(n \times n)$ matrix and Σ a rectangular $(m \times n)$ matrix, diagonal in the sense described below. The columns of the orthogonal matrices U and V are called the (left and right) singular vectors of the matrix A. If we denote $r = \text{rank}(A)$, the maximum number of linearly independent rows (or columns) of A, then the matrix Σ has all entries zero except the first r entries on the main diagonal, which are positive and are called the singular values of A. The convention is to have the singular values arranged in decreasing order: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. We extend the notation

$$\Sigma = \text{diag}_{m \times n}\{\sigma_1, \ldots \sigma_r\}$$

for this kind of 'diagonal' rectangular $(m \times n)$ matrix. Such SVD is not unique, but the singular values (hence Σ) are.

Construction of the SVD

First, assuming the SVD already exists, we investigate what kind of matrices V and Σ must represent. From $A = U\Sigma V^T$ we compute $A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = (V\Sigma^T U^T)(U\Sigma V^T) = V(\Sigma^T\Sigma)V^T$. Therefore V and $\Sigma^T\Sigma$ are two matrices that diagonalize the matrix A^TA. We now recognize that this is always possible precisely because the matrix A^TA is indeed symmetric. We could make the same argument about the matrix AA^T, but it turns out that the relationship between U and V is much more intimate. In fact $AV = U\Sigma$ means that each column vector of V is mapped into a scalar multiple of a (column) vector of U. With these observations, we now detail the recipe for constructing an SVD for A.

Starting with the matrix A, compute A^TA, which is a $n \times n$ symmetric, semi-positive definite (i.e. with nonnegative eigenvalues) matrix with the same rank as A: rank(A^TA)=rank(A)=r. Being symmetric, is it diagonalizable, hence one can find an orthonormal basis of eigenvectors for A^TA, denoted $V_1, V_2, \ldots, V_n$. Let V be the $(n \times n)$ eigenvector matrix with columns V_j. The diagonal matrix D will have only nonnegative diagonal entries, since A^TA is semi-positive definite. Upon arranging them in decreasing order (by eventually permuting the columns of V), and denoting by σ_i the square root of each positive eigenvalue, we have

$$D = \text{ diag }_{n\times n}\{\sigma_1^2, \ldots \sigma_r^2, 0, \ldots, 0\}.$$

Note that there are precisely $n - r$ zeros in D since that is the dimension of the kernel of A^TA, as given by the dimension theorem. Now assemble the rectangular $(m \times n)$ matrix Σ using the entries $\sigma_1, \ldots, \sigma_r$ on the first r entries of the main diagonal, the rest of Σ entries being zero. Clearly $\Sigma^T\Sigma = D$.

The final step of the decomposition is as follows: if $V_1, \ldots, V_r$ are the first r columns of V, then, for $j = 1 \ldots r$, $|AV_j|^2 = (AV_j)^T(AV_j) = V_j^T(A^TA)V_j = V_j^T(V\Sigma^T\Sigma V^T)V_j = (V_j^TV)D(V^TV_j) = \sigma_j^2$, so the vector AV_j has length equal to σ_j. We can now define the vectors $U_1, \ldots U_r$, where U_j is the *unit* vector obtained by normalizing AV_j: $U_j = \frac{1}{|AV_j|}AV_j = \frac{1}{\sigma_j}AV_j$, or

$$AV_j = \sigma_j U_j, \quad j = 1, \ldots r$$

The set of vectors $\{U_1, \ldots U_r\}$ forms an orthonormal basis for the column space of A. We can extend this set to an orthonormal basis of $\mathbb{R}^m$ by completing it with vectors $U_{r+1}, \ldots U_m$ from the left nullspace of A. Since $V_{r+1}, \ldots V_n$ belong to the kernel of A, we have just established the relation

$$AV = U\Sigma$$

which yields $A = U\Sigma V^{-1} = U\Sigma V^T$, the desired decomposition (4).

The SVD construction presented above, remarkably simple, makes the result even more beautiful and powerful. SVD is considered by many (see (Strang 2009)) to be one of the jewels of linear algebra. The simplicity of this makes the numerical implementation of the SVD straightforward.

A final word about SVD for symmetric matrices: if the matrix A is symmetric (hence square) and has eigenvalues $\lambda_1, \ldots, \lambda_n$ the singular values are precisely the positive parts of the eigenvalues of A, ordered in decreasing order, since the construction above gives σ_i^2 to be the eigenvalues of $A^TA = A^2$. IN this sense, SVD can be regarded as a generalization of the diagonalization of symmetric matrices.

Computational aspects

The computation of eigenvalues and eigenvectors can be performed via standard numerical algorithms, or, in case of large matrices, using more sophisticated algorithms, depending on the structure of the matrix (e.g. sparse, random etc). A standard algorithm for eigenvalue computation is the QR iteration method, and one can find its implementation in computational platforms such as MATLAB, (Moler 2004), via the `eig` built-in function. The computation of the SVD of the matrix A is, at least in theory, equivalent to the computation of eigenvalues and eigenvectors for the matrix $A^T A$. An effective implementation of such an algorithm in MATLAB is the `svd` function. **Optimization of such computations in the case of large sparse matrices can be found in (Saad 2011).**

Applications

Matrix eigenvalue and singular value computations are essential in a wide range of applications, from structural dynamics, power networks, image processing and data mining, stability and control in dynamical systems, to social network analysis andy crowd dynamics, just to name a few. Below we detail just a few instances where matrix analysis is used in applications.

Image Compression

When a large matrix is used to encode some information (such as an image), then compression of this data can be done using SVD. More specifically, SVD finds low rank approximations of the original matrix which preserve the 'essential' information. If $A = U\Sigma V^T$

then by keeping only the first k largest singular values ($k < r$) and the matrices U and V to the corresponding k singular vectors can construct the rank-k matrix

$$A_k = U_k \Sigma_k V_k^T$$

which approximates the matrix A. Image compression can be effective especially when the $(k+1)^{th}$ singular value is significantly smaller than the first k ones.

Data Mining

Linear algebra tools are ubiquitous in data mining applications. For example, face recognition (or handwriting recognition) are classical examples where eigenvalue computation and SVD can be applied. From a large set of data one identifies a few representative ones, called eigenfaces, and then projects all the other ones. See (Kokiopoulou 2011) for a good overview of such dimension reduction methods.

One particularly important tool in analyzing multidimensional data is the Principal Component Analysis (PCA). This is a method which extracts the 'essential' structure of the data set in the form of a lower dimensional subspace, spanned by so-called principal components. **See e.g. (Jolliffe 2002) or (Abdi 2010).** Principal components have a particular ordering – each principal component points in the direction of maximal variance that is orthogonal to each of the previous principal components. In this way, each principal component accounts for the maximal possible amount of variance, ignoring the variance already accounted for by the previous principal components. These directions turn out to be precisely the eigenvectors of $A^T A$. See (Johnson 2007) for a detailed description of the PCA, its geometric representations and its application in multivariate data analysis.

Below are illustrations of a data set before and after the PCA analysis. Data was collected from 75 universities in the United States (MUP 2012) and consisted of 8 variables: level of total research funds, federal research funds, endowment assets, annual giving, number of National Academy of Science (NAS) members, number of faculty awards and number of doctorates and postdocs. An 8-dimensional data point corresponds to each University in the study.

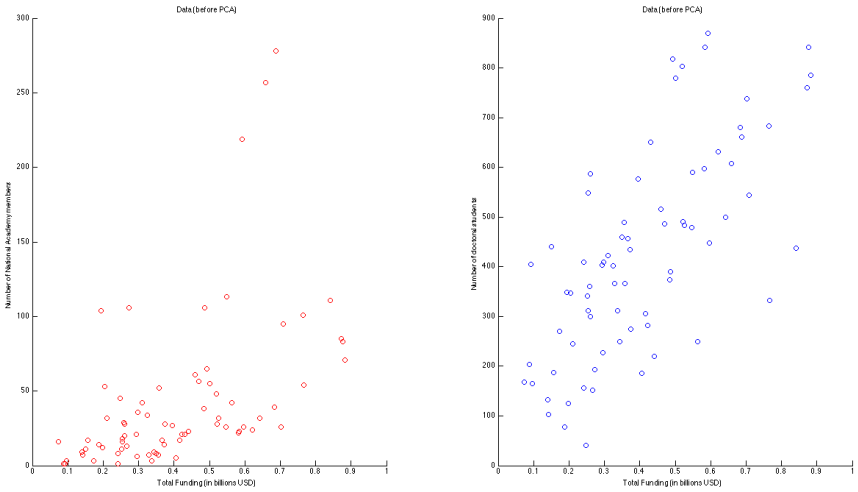

Figure 1. Partial representation of the 8-dim university data set, before PCA, using only pairs of the variables: number of NAS members (left) and number of doctoral students (right) vs. the total research funding.

The PCA computes the eigenvalues (principal values) and eigenvectors (principal components) of the correlation matrix formed by the renormalized (to zero mean) variables. Then each variable (column of A) is projected to the first few principal components. The first few principal components for the entire data set are plotted below.

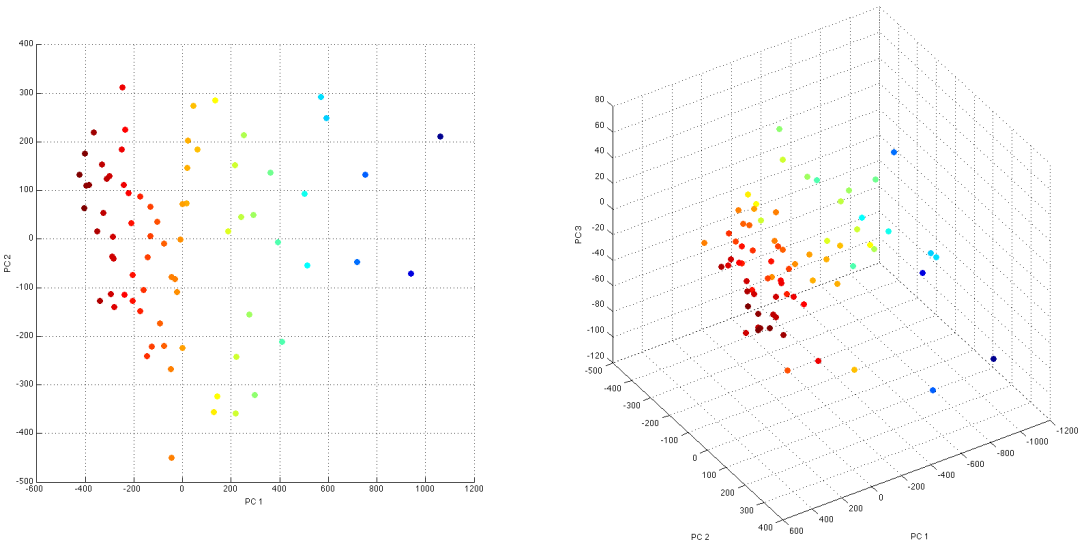

Figure 2. Projection of the university data set onto the first two (left) and three (right) principal component directions.

Here each data point is color coded, depending on the magnitude of the first principal component. Visualizing such high dimensional data using projections to lower dimensional space is useful in identifying directions of maximal variation in the data, which, in the case of the university data example, could be used in university ranking.

Network Analysis

The study of networks has become central in many scientific fields; see e.g. (Easley 2010). The mathematical representation of a network often associates certain matrices to the network, such as the adjacency matrix or the graph Laplacian (Neuman 2010). Understanding the type of information gained by matrix analysis is in general the crux of the application. In the case of large networks, additional computational challenges must be mitigated.

The adjacency matrix $A = (a_{ij})$ of a network consists of 1's and 0's, with $a_{ij} = 1$ if the i^{th} node is connected with the j^{th} node and $a_{ij} = 0$ otherwise. For undirected

networks, the adjacency matrix is symmetric, while for directed networks, it is not symmetric. The eigenvector corresponding to the largest eigenvalue can be used to define the node centrality, that is the 'importance' of each node in the network. This is not simply the node with the higher number of 'connections', but it also takes into account the 'importance' of connected nodes. A slightly more involved notion of centrality has been used in establishing the Google's PageRank algorithm.

One interesting application of the eigenvalue computation is related to the graph partitioning problem. The second-highest eigenvalue gives a measure of how easy it is to partition the network in two (comparable in size) while keeping the number of edges between sides to a minimum. This can be generalized to multiple partitions. In social networks, this can be used to determine the community structure (Neuman 2010).

Interpreting the SVD can present challenges depending on the application. While the singular vector corresponding to the leading singular value usually can be thought as the principal direction in which the data is organized, it is often the case that the subsequent singular values cannot always be associated with a specific property of the network in question, other than that they are orthogonal to previous ones. Instead, interpreting the SVD must be inferred from the application at hand (Mahoney 2011).

Cross-references

Related Essays: 88, 136, 152, 153, 332 [NEED TO HAVE UPDATED TOC!!!!]

References

Abdi 2010. Abdi H and Williams L J (2010), Principal component analysis. In WIREs Comp Stat, 2: 433–459

MUP 2012. The Center for Measuring University Performance, http://mup.asu.edu/ accessed Sept 2013

Costa 2010. Costa L da F, Villas Boas P R, Silva F N and Rodrigues F A (2010) A pattern recognition approach to complex networks. In: Journal of Statistical Mechanics. Theory and Experiments.

Easley 2010. Easley D, Kleinberg J (2010) Networks, Crowds, and Markets, Cambridge Univ Press, New York.

Jolliffe 2002. Jolliffe I T (2002) Principal Component Analysis, Springer, New York.

Kokiopoulou 2011. Kokiopoulou E, Chen J, Saad Y (2011) Trace optimization and eigenproblems in dimension reduction methods. In: Numerical Linear Algebra with Applications (NLAA), 18(3):565–602

Johnson 2007. Johnson R and Wichern D (2007) Applied Multivariate Statistical Analysis, sixth edition, Pearson

Mahoney 2011. Mahoney M (2011) Randomized Algorithms for Matrices and Data, Foundations and Trends in Machine Learning: Vol. 3: No 2, pp 123-224. http://dx.doi.org/10.1561/2200000035

Moler 2004. Moler C (2004) Numerical Computing with MATLAB, SIAM Philadelphia, freely available at http://www.mathworks.it/moler/, accessed April 2013.

Neuman 2010. Newman M (2010) Networks: An Introduction. Oxford, UK: Oxford University Press.

Saad 2011. Saad Y (2011) Numerical Methods for Large Eigenvalue Problems, 2nd edition, Classics in Applied Mathematics, SIAM, Philadelphia; freely available at http://www.siam.org/books/cl66 , accessed April 2013.

Strang 2009. Strang G (2009) Linear Algebra and Applications, 4th edition, Brooks Cole, NY