Singular Value Decomposition

Kenichi Kanatani
Professor Emeritus, Okayama University, Japan

Related Concepts

- Spectral Decomposition
- Pseudoinverse
- Karhunen–Loéve Expansion
- Principal Component Analysis
- Factorization

Definition

Spectral decomposition is an expression of a symmetric matrix in terms of its eigenvalues and eigen vectors, while singular value decomposition is a similar expression of a nonzero rectangular matrix in terms of its singular values and singular vectors.

Background

A linear mapping from $\mathcal{R}^n$ to $\mathcal{R}^m$ is represented by an $m \times n$ matrix $\boldsymbol{A}$. It is determined by the images $\boldsymbol{a}_1$, ..., $\boldsymbol{a}_n \in \mathcal{R}^m$ of an orthonormal basis $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ of $\mathcal{R}^n$ in the form

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{a}_i \boldsymbol{u}_i^\top. \tag{1}$$

From the orthogonality $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \delta_{ij}$ (Kronecker delta), where $\langle \cdot\,, \cdot \rangle$ denotes the inner product of vectors, we can confirm that $\boldsymbol{A}\boldsymbol{u}_i = \boldsymbol{a}_i$ holds indeed.

For an $n \times n$ symmetric matrix $\boldsymbol{A}$, there exist a real value λ, called *eigenvalue*, and a nonzero vector $\boldsymbol{u}$ ($\neq \boldsymbol{0}$), called *eigenvector*, such that

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}. \tag{2}$$

An $n \times n$ symmetric matrix has n eigenvalues λ_1, ..., λ_n (with possible overlaps), and the corresponding eigenvectors $\boldsymbol{u}_1$, ..., $\boldsymbol{u}_n$ can be chosen to be mutually orthogonal unit vectors. Hence, $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ serve as an orthonormal basis of $\mathcal{R}^n$. Eigenvalues and eigenvectors are easily computed using mathematical software [1, 2].

The relationship $\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$, $i = 1$, ..., n, implies that $\boldsymbol{A}$ maps an orthonormal basis $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ to $\lambda_1 \boldsymbol{u}_1$, ..., $\lambda_n \boldsymbol{u}_n$. Letting $\boldsymbol{a}_i = \lambda_i \boldsymbol{u}_i$ in (1), we can write the matrix $\boldsymbol{A}$ in the form

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top, \tag{3}$$

which is called the *spectral* (or *eigenvalue*) *decomposition* of $\boldsymbol{A}$.

Theory

For a nonzero $m \times n$ matrix $\boldsymbol{A}$ ($\neq \boldsymbol{O}$), there exists a positive value σ (> 0), called *singular value*, a nonzero m-dimensional vector $\boldsymbol{u}$ ($\neq \boldsymbol{0}$), called *left singular vector*, and a nonzero n-dimensional vector $\boldsymbol{v}$ ($\neq \boldsymbol{0}$), called *right singular vector*, such that

$$\boldsymbol{A}\boldsymbol{v} = \sigma \boldsymbol{u}, \qquad \boldsymbol{A}^\top \boldsymbol{u} = \sigma \boldsymbol{v}. \tag{4}$$

The left and right singular vectors are referred to simply as *singular vectors*.

Multiplying the first and the second equations of (4) by $\boldsymbol{A}^\top$ and $\boldsymbol{A}$, respectively, we can see that

$$\boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{u} = \sigma^2 \boldsymbol{u}, \qquad \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{v} = \sigma^2 \boldsymbol{v}. \tag{5}$$

Namely, $\boldsymbol{u}$ and $\boldsymbol{v}$ are, respectively, the eigenvectors of the $m \times m$ symmetric matrix $\boldsymbol{A}\boldsymbol{A}^\top$ and of the $n \times n$ symmetric matrix $\boldsymbol{A}^\top \boldsymbol{A}$ for eigenvalue σ^2. Hence, the number of singular values of $\boldsymbol{A}$ equals the number of nonzero eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{A}^\top \boldsymbol{A}$, which is equal to the rank r of $\boldsymbol{A}$. Singular values and singular vectors are easily computed using mathematical software [1, 2].

Since the singular vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are eigenvectors of symmetric matrices, the singular vectors $\{\boldsymbol{u}_i\}$ and $\{\boldsymbol{v}_i\}$, $i = 1$, ..., r, can be chosen as orthonormal sets. The set $\{\boldsymbol{v}_i\}$, $i = 1$, ..., r, can be extended to an orthonormal basis $\{\boldsymbol{v}_i\}$, $i = 1$, ..., n, of $\mathcal{R}^m$, in such a way that $\boldsymbol{v}_i$, $i = r+1$, ..., n, are the null vectors, i.e., eigenvectors for eigenvalue 0, of $\boldsymbol{A}^\top \boldsymbol{A}$. Hence,

$$\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}, \qquad i = r+1, ..., n. \tag{6}$$

This and (4) imply that $\boldsymbol{A}$ maps an orthonormal basis $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_m\}$ to $\sigma_1 \boldsymbol{u}_1$, ..., $\sigma_r \boldsymbol{u}_r$, $\boldsymbol{0}$, ..., $\boldsymbol{0}$. Hence, from (1), the matrix $\boldsymbol{A}$ is written as

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top, \tag{7}$$

which is called the *singular value decomposition* (*SVD*) of $\boldsymbol{A}$.

Pseudoinverse

If an $m \times n$ matrix $\boldsymbol{A}$ ($\neq \boldsymbol{O}$) has the singular value decomposition (7), its *pseudoinverse* (or *generalized inverse*) of *Moore-Penrose type* is defined by

$$\boldsymbol{A}^- = \sum_{i=1}^{r} \frac{\boldsymbol{v}_i \boldsymbol{u}_i^\top}{\sigma_i}. \tag{8}$$

Some authors write $\boldsymbol{A}^-$ for general (not necessarily of Moore-Penrose type) pseudoinverse and specifically write $\boldsymbol{A}^+$ for the Moore-Penrose type to distinguish it from others.

Note that inverse is defined only for nonsingular (hence square) matrices while pseudoinverse is defined for all nonzero (generally rectangular) matrices. The inverse of a nonsingular matrix is defined so that its product with the original matrix is the identity. This is not necessarily so for pseudoinverse. From (7) and (8), we can see that

$$\boldsymbol{A}\boldsymbol{A}^- = \sum_{i=1}^{r} \boldsymbol{u}_i \boldsymbol{u}_i^\top \quad (\triangleq \boldsymbol{P}_{\mathcal{U}}), \tag{9}$$

$$\boldsymbol{A}^-\boldsymbol{A} = \sum_{i=1}^{r} \boldsymbol{v}_i \boldsymbol{v}_i^\top \quad (\triangleq \boldsymbol{P}_{\mathcal{V}}). \tag{10}$$

The matrix $\boldsymbol{P}_{\mathcal{U}}$ represents projection onto the subspace $\mathcal{U}$ spanned by $\boldsymbol{u}_1$, ..., $\boldsymbol{u}_r$. In fact, the orthogonality $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \delta_{ij}$ implies that $\boldsymbol{P}_{\mathcal{U}}\boldsymbol{u} = \boldsymbol{u}$ for $\boldsymbol{u} \in \mathcal{U}$ and $\boldsymbol{P}_{\mathcal{U}}\boldsymbol{u} = \boldsymbol{0}$ for $\boldsymbol{u} \perp \mathcal{U}$. From (7), we see that $\mathcal{U}$ coincides with the subspace spanned by the columns of $\boldsymbol{A}$. Similarly, the matrix $\boldsymbol{P}_{\mathcal{V}}$ represents projection onto the subspace $\mathcal{V}$ spanned by the rows of $\boldsymbol{A}$. Thus, the product of a matrix $\boldsymbol{A}$ with its pseudoinverse $\boldsymbol{A}^-$ is the projection onto the subspace spanned by the columns or the rows of $\boldsymbol{A}$. Since the columns and rows of a nonsingular matrix span the entire space, for which the identity is the projection onto it, pseudoinverse is a natural extension of the inverse to arbitrary rectangular matrices.

Since the columns of $\boldsymbol{A}$ span the subspace $\mathcal{U}$ and the rows span $\mathcal{V}$, the following identities hold:

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}, \qquad \boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^- = \boldsymbol{A}^-. \tag{11}$$

Historically, the first equation is the definition the general (not necessarily of Moore-Penrose type) pseudoinverse, and the Moore-Penrose type is obtained by requiring the second equation (called *reflexivity*) and the condition that $\boldsymbol{A}\boldsymbol{A}^-$ and $\boldsymbol{A}^-\boldsymbol{A}$ be symmetric.

Since pseudoinverse is defined for any nonzero matrices, it can be defined for vectors, regarded as $n \times 1$ and $1 \times n$ matrices. For a vector $\boldsymbol{a}$, the definition of pseudoinverse implies

$$\boldsymbol{a}^- = \frac{\boldsymbol{a}^\top}{\|\boldsymbol{a}\|^2}, \qquad (\boldsymbol{a}^\top)^- = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|^2}. \tag{12}$$

Consider a linear equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{13}$$

where $\boldsymbol{A}$ is a nonzero $m \times n$ matrix, $\boldsymbol{x}$ is an n-dimensional unknown vector, and $\boldsymbol{b}$ is a given m-dimensional vector. The expression

$$\boldsymbol{x} = \boldsymbol{A}^-\boldsymbol{b} \tag{14}$$

gives a *least-squares solution* such that (i) it minimizes $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$ over all $\boldsymbol{x}$ and (ii) if the minimum is not unique, it returns $\boldsymbol{x}$ for which $\|\boldsymbol{x}\|^2$ is the smallest among them. This procedure is an essential component of multiview 3D reconstruction, using redundant geometric constraints in the presence of noise [3].

Subspace Fitting

Given N points $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ in $\mathcal{R}^n$, we consider the problem of finding an r-dimensional subspace ($r < N$) that is the closest to them, where the closeness is measured by the sum of square distances. We assume that the coordinate system is translated so that the origin O coincides with the centroid of the N points. Let

$$\boldsymbol{S} = \sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha^\top. \tag{15}$$

This matrix is called by many different names such as the "moment matrix" and the "scatter matrix" (both from physics). Here, let us call it, for convenience, the *covariance matrix*, a term borrowed from statistics.

The form of (15) implies that it is a positive semidefinite symmetric matrix, having nonnegative eigenvalues. Let σ_i^2 be its eigenvalues, and $\boldsymbol{u}_i$ the corresponding unit eigenvectors, $i = 1$, ..., n. The spectral decomposition of $\boldsymbol{S}$ has the form

$$\boldsymbol{S} = \sum_{i=1}^{n} \sigma_i^2 \boldsymbol{u}_i \boldsymbol{u}_i^\top. \tag{16}$$

We can show that $\boldsymbol{u}_1$ is the direction of the line from which the sum of square distances of the N points is the smallest; $\boldsymbol{u}_2$ is the direction of the line from which the sum of square distances of the N points is the smallest among all directions orthogonal to $\boldsymbol{u}_1$; $\boldsymbol{u}_3$ is the direction of the line from which the sum of square distances of the N points is the smallest among all directions orthogonal to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, and so on. Thus, it is concluded that the r-dimensional subspace that best approximates the N points is spanned by $\boldsymbol{u}_1$, ..., $\boldsymbol{u}_r$. They are called the *principal directions* of the N points.

Hierarchical construction of such subspaces of dimensions $r = 1, 2, \ldots$ is called the *Karhunen–Loéve* (*KL*) *expansion* in the domain of signal processing, expanding individual signals with respect to $\{\boldsymbol{u}_i\}$. The efficiency of transmission and display of the signal improves by omitting those basis vectors with small contributions. In statistics, this scheme is called the *principal component analysis* (*PCA*), used for extracting a small number of statistics that characterize the data well.

For most vision applications [4], however, *we need not compute the covariance matrix* $\boldsymbol{S}$. In fact, let

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_N \end{pmatrix} \tag{17}$$

be the $n \times N$ matrix consisting of the data vectors $\boldsymbol{x}_\alpha$, $\alpha = 1$, ..., N, as its columns. The covariance matrix $\boldsymbol{S}$ of (16) equals

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\top. \tag{18}$$

The singular value decomposition of $\boldsymbol{X}$ has the form

$$\boldsymbol{X} = \sum_{i=1}^{n} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top, \tag{19}$$

because, as seen from (5), the eigenvalues of $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\top$ equals the squares of the singular values of $\boldsymbol{X}$; $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the eigenvectors of $\boldsymbol{X}\boldsymbol{X}^\top$ and $\boldsymbol{X}^\top\boldsymbol{X}$, respectively. Hence, the basis of the fitted r-dimensional subspace is given by the singular vectors $\boldsymbol{u}_1$, ..., $\boldsymbol{u}_r$ of $\boldsymbol{X}$. This approach considerably reduces the computational complexity, since we need not compute the covariance matrix $\boldsymbol{S}$, which requires $O(n^2N)$ operations, and the complexity of eigenanalysis is $O(n^3)$, while for singular value decomposition it is $O(n^2N)$ for $N \geq n$ and $O(nN^2)$ for $n \geq N$ [1], i.e., the complexity ls *linear* in the length of the "shorter side" of the matrix.

Exploiting this fact, we can see that 3D reconstruction, for example, from hundreds of thousands of data points, which would require several hours of computation using spectral decomposition, can reduce to several seconds using singlar value decomposition [3].

Matrix Product Representation

The spectral decomposition (3) is written as

$$\begin{aligned}
\boldsymbol{A} &= \begin{pmatrix} \lambda_1\boldsymbol{u}_1 & \cdots & \lambda_n\boldsymbol{u}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1^\top \\ \vdots \\ \boldsymbol{u}_n^\top \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1^\top \\ \vdots \\ \boldsymbol{u}_n^\top \end{pmatrix} \\
&= \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top,
\end{aligned} \tag{20}$$

where $\boldsymbol{U}$ is the $n \times n$ matrix consisting of the eigenvectors $\boldsymbol{u}_i$ as its columns and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_n)$. Multiplying the above equation by $\boldsymbol{U}$ from right and $\boldsymbol{U}^\top$ from left, we obtain $\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}$, which is known as *diagonalization* of the symmetric matrix $\boldsymbol{A}$.

Using the same rewriting as (20), we can express the singular value decomposition (7) in the form

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top, \tag{21}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the $m \times r$ and $n \times r$ matrices consisting of the singular vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ as their columns, respectively, and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, ..., \sigma_r)$. Then, the pseudoinverse (8) is written as

$$\boldsymbol{A}^- = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^\top. \tag{22}$$

Application

The singular value decomposition in the form of (21) is utilized in the method of *factorization* for 3D reconstruction from images [4, 5]. The singular value decomposition also plays a central role in 3D vision computation problems including optimal fundamental matrix estimation, optimal rotation estimation, and multiview 3D reconstruction computation [3].

References

[1] Golub, G.H., Van Loan, C.F. (2012) Matrix Computations. 4th ed., Johns Hopkins Univ. Press, Baltimore

[2] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. (2004). Numerical Recipes: The Art of Scientific Computing. 3rd ed. Cambridge Univ. Press, Cambridge

[3] Kanatani, K., Sugaya, Y., Kanazawa, K. (2016) Guide to 3D Vision Computation: Geometric Analysis and Implementation. Springer, Switzerland

[4] Hartley, R., Zisserman, A. (2003) Multiple View Geometry in Computer Vision. 2nd ed., Cambridge Univ. Press, Cambridge

[5] Tomasi, C., Kanade, T. (1992) Shape and motion from image streams under orthography—A factorization method. Int. J. Comput. Vis. 9(2): 137–154