Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD)

Hervé Abdi[1]

1 Overview

The singular value decomposition (SVD) is a generalization of the eigen-decomposition which can be used to analyze rectangular matrices (the eigen-decomposition is defined only for squared matrices). By analogy with the eigen-decomposition, which decomposes a matrix into two simple matrices, the main idea of the SVD is to decompose a rectangular matrix into three simple matrices: Two orthogonal matrices and one diagonal matrix.

Because it gives a least square estimate of a given matrix by a lower rank matrix of same dimensions, the SVD is equivalent to principal component analysis (PCA) and metric multidimensional

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

scaling (MDS) and is therefore an essential tool for multivariate analysis. The generalized SVD (GSVD) decomposes a rectangular matrix and takes into account constraints imposed on the rows and the columns of the matrix. The GSVD gives a weighted generalized least square estimate of a given matrix by a lower rank matrix and therefore, with an adequate choice of the constraints, the GSVD implements all linear multivariate techniques (*e.g.,* canonical correlation, linear discriminant analysis, correspondence analysis, PLS-regression).

2 Definitions and notations

Recall that a *positive semi-definite* matrix can be obtained as the product of a matrix by its transpose. This matrix is obviously square and symmetric, but also (and this is less obvious) its eigenvalues are all positive or null, and the eigenvectors corresponding to different eigenvalues are pairwise orthogonal. Let $\mathbf{X}$ be a positive semi-definite, its eigen-decomposition is expressed as

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathsf{T}} , \tag{1}$$

with $\mathbf{U}$ being an orthonormal matrix (*i.e.,* $\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I}$) and $\boldsymbol{\Lambda}$ being a diagonal matrix containing the eigenvalues of $\mathbf{X}$.

The SVD uses the eigen-decomposition of a positive semi-definite matrix in order to derive a similar decomposition applicable to all rectangular matrices composed of real numbers. The idea here is to decompose any matrix into three simple matrices, two orthonormal matrices and one diagonal matrix. When applied to a *positive semi-definite* matrix, the SVD is equivalent to the eigen-decomposition.

Formally, if $\mathbf{A}$ is a rectangular matrix, its SVD decomposes it as:

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}} , \tag{2}$$

with:

- $\mathbf{P}$: the (normalized) eigenvectors of the matrix $\mathbf{A}\mathbf{A}^{\mathsf{T}}$ (*i.e.,* $\mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{I}$). The columns of $\mathbf{P}$ are called the *left singular vectors* of $\mathbf{A}$.

- $\mathbf{Q}$: the (normalized) eigenvectors of the matrix $\mathbf{A}^\top\mathbf{A}$ (*i.e.,* $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$). The columns of $\mathbf{Q}$ are called the *right singular vectors* of $\mathbf{A}$.

- $\boldsymbol{\Delta}$: the diagonal matrix of the *singular values*, $\boldsymbol{\Delta} = \boldsymbol{\Lambda}^{\frac{1}{2}}$ with $\boldsymbol{\Lambda}$ being the diagonal matrix of the eigenvalues of matrix $\mathbf{A}\mathbf{A}^\top$ and of the matrix $\mathbf{A}^\top\mathbf{A}$ (they are the same).

The SVD is a consequence of the eigen-decomposition of a positive semi-definite matrix. This can be shown by considering the eigen-decomposition of the two positive semi-definite matrices that can be obtained from $\mathbf{A}$: namely $\mathbf{A}\mathbf{A}^\top$ and $\mathbf{A}^\top\mathbf{A}$. If we express these matrices in terms of the SVD of $\mathbf{A}$, we obtain the following equations:

$$\mathbf{A}\mathbf{A}^\top = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^\top\mathbf{Q}\boldsymbol{\Delta}\mathbf{P}^\top = \mathbf{P}\boldsymbol{\Delta}^2\mathbf{P}^\top = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^\top , \tag{3}$$

and

$$\mathbf{A}^\top\mathbf{A} = \mathbf{Q}\boldsymbol{\Delta}\mathbf{P}^\top\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^\top = \mathbf{Q}\boldsymbol{\Delta}^2\mathbf{Q}^\top = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top . \tag{4}$$

This shows that $\boldsymbol{\Delta}$ is the square root of $\boldsymbol{\Lambda}$, that $\mathbf{P}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^\top$, and that $\mathbf{Q}$ are the eigenvectors of $\mathbf{A}^\top\mathbf{A}$.

For example, the matrix:

$$\mathbf{A} = \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} \tag{5}$$

can be expressed as:

$$\begin{aligned} \mathbf{A} &= \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^\top \\ &= \begin{bmatrix} 0.8165 & 0 \\ -0.4082 & -0.7071 \\ -0.4082 & 0.7071 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \\ &= \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} . \end{aligned} \tag{6}$$

We can check that:

$$\mathbf{A}\mathbf{A}^{\mathsf{T}} = \begin{bmatrix} 0.8165 & 0 \\ -0.4082 & -0.7071 \\ -0.4082 & 0.7071 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 1^2 \end{bmatrix} \begin{bmatrix} 0.8165 & -0.4082 & -0.4082 \\ 0 & -0.7071 & 0.7071 \end{bmatrix}$$

$$= \begin{bmatrix} 2.6667 & -1.3333 & -1.3333 \\ -1.3333 & 1.1667 & 0.1667 \\ -1.3333 & 0.1667 & 1.1667 \end{bmatrix} \tag{7}$$

and that:

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 1^2 \end{bmatrix} \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}$$

$$= \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix} . \tag{8}$$

2.1 Technical note: Agreement between signs

Singular vectors come in pairs made of one left and one right singular vectors corresponding to the same singular value. They could be computed separately or as a pair. Equation 2 requires computing the eigen-decomposition of two matrices. Rewriting this equation shows that it is possible, in fact, to compute only one eigen-decomposition. As an additional bonus, computing only one eigen-decomposition prevents a problem which can arise when the singular vectors are obtained from two separate eigen-decompositions. This problem follows from the fact that the eigenvectors of a matrix are determined up to a multiplication by -1, but that singular vectors being pairs of eigenvectors need to have compatible parities. Therefore, when computed as eigenvectors, a pair of singular vectors can fail to reconstruct the original matrix because of this parity problem.

This problem is illustrated by the following example: The matrix

$$\mathbf{A} = \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} \tag{9}$$

can be decomposed in two equivalent ways:

$$\begin{aligned}
\mathbf{A} &= \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}} \\
&= \begin{bmatrix} 0.8165 & 0 \\ -0.4083 & -0.7071 \\ -0.4083 & 0.7071 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \\
&= \begin{bmatrix} -0.8165 & 0 \\ 0.4083 & 0.7071 \\ 0.4083 & 0.7071 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{bmatrix} \\
&= \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} .
\end{aligned} \tag{10}$$

But when the parity of the singular vectors does not match, the SVD will fail to reconstruct the original matrix as illustrated by

$$\begin{aligned}
\mathbf{A} &\neq \begin{bmatrix} -0.8165 & 0 \\ 0.4083 & 0.7071 \\ 0.4083 & 0.7071 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \\
&= \begin{bmatrix} -1.1547 & -1.1547 \\ 0.0774 & 1.0774 \\ 1.0774 & 0.0774 \end{bmatrix} .
\end{aligned} \tag{11}$$

By computing only one matrix of singular vectors, we can rewrite Equation 2 in a manner that expresses that one matrix of singular vectors can be obtained from the other:

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}} \quad \Longleftrightarrow \quad \mathbf{P} = \mathbf{A}\mathbf{Q}\boldsymbol{\Delta}^{-1} . \tag{12}$$

For example:

$$
\begin{aligned}
\mathbf{P} &= \mathbf{A}\mathbf{Q}\boldsymbol{\Delta}^{-1} \\
&= \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0.8165 & 0 \\ -0.4082 & -0.7071 \\ -0.4082 & 0.7071 \end{bmatrix} .
\end{aligned} \tag{13}
$$

3 Generalized singular value decomposition

For a given $I \times J$ matrix $\mathbf{A}$, generalizing the singular value decomposition, involves using two positive definite square matrices with size $I \times I$ and $J \times J$ respectively. These two matrices express constraints imposed respectively on the rows and the columns of $\mathbf{A}$. Formally, if $\mathbf{M}$ is the $I \times I$ matrix expressing the constraints for the rows of $\mathbf{A}$ and $\mathbf{W}$ the $J \times J$ matrix of the constraints for the columns of $\mathbf{A}$. The matrix $\mathbf{A}$ is now decomposed into:

$$
\mathbf{A} = \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{V}}^{\mathsf{T}} \qquad \text{with: } \widetilde{\mathbf{U}}^{\mathsf{T}}\mathbf{M}\widetilde{\mathbf{U}} = \widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{W}\widetilde{\mathbf{V}} = \mathbf{I} . \tag{14}
$$

In other words, the generalized singular vectors are orthogonal under the constraints imposed by $\mathbf{M}$ and $\mathbf{W}$.

This decomposition is obtained as a result of the standard singular value decomposition. We begin by defining the matrix $\widetilde{\mathbf{A}}$ as:

$$
\widetilde{\mathbf{A}} = \mathbf{M}^{\frac{1}{2}}\mathbf{A}\mathbf{W}^{\frac{1}{2}} \iff \mathbf{A} = \mathbf{M}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\mathbf{W}^{-\frac{1}{2}} . \tag{15}
$$

We then compute the standard singular value decomposition as $\widetilde{\mathbf{A}}$ as:

$$
\widetilde{\mathbf{A}} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}} \qquad \text{with: } \mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I} . \tag{16}
$$

The matrices of the generalized eigenvectors are obtained as:

$$
\widetilde{\mathbf{U}} = \mathbf{M}^{-\frac{1}{2}}\mathbf{P} \quad \text{and} \quad \widetilde{\mathbf{V}} = \mathbf{W}^{-\frac{1}{2}}\mathbf{Q} . \tag{17}
$$

The diagonal matrix of singular values is simply equal to the matrix of singular values of $\widetilde{\mathbf{A}}$:

$$\widetilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta} \tag{18}$$

We verify that:

$$\mathbf{A} = \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{V}}^{\mathsf{T}}$$

by substitution:

$$\begin{aligned}\mathbf{A} &= \mathbf{M}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\mathbf{W}^{-\frac{1}{2}}\\ &= \mathbf{M}^{-\frac{1}{2}}\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}}\mathbf{W}^{-\frac{1}{2}}\\ &= \widetilde{\mathbf{U}}\boldsymbol{\Delta}\widetilde{\mathbf{V}}^{\mathsf{T}} \qquad \text{(from Equation 17)}\ .\end{aligned} \tag{19}$$

To show that Condition 14 holds, suffice to show that:

$$\widetilde{\mathbf{U}}^{\mathsf{T}}\mathbf{M}\widetilde{\mathbf{U}} = \mathbf{P}^{\mathsf{T}}\mathbf{M}^{-\frac{1}{2}}\mathbf{M}\mathbf{M}^{-\frac{1}{2}}\mathbf{P} = \mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{I} \tag{20}$$

and

$$\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{W}\widetilde{\mathbf{V}} = \mathbf{Q}^{\mathsf{T}}\mathbf{W}^{-\frac{1}{2}}\mathbf{W}\mathbf{W}^{-\frac{1}{2}}\mathbf{Q} = \mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I}\ . \tag{21}$$

4 Mathematical properties

It can be shown that (see, *e.g.,* Strang, 2003; Abdi & Valentin 2006) that the SVD has the important property of giving an optimal approximation of a matrix by another matrix of smaller rank. In particular, the SVD gives the best approximation, in a least square sense, of any rectangular matrix by another rectangular of same dimensions, but smaller rank.

Precisely, if $\mathbf{A}$ is an $I \times J$ matrix of rank L (*i.e.,* $\mathbf{A}$ contains L singular values that are not zero), we denote by $\mathbf{P}_{[K]}$ (respectively $\mathbf{Q}_{[K]}$, $\boldsymbol{\Delta}_{[K]}$) the matrix made of the first K columns of $\mathbf{P}$ (respectively $\mathbf{Q}$,

$\mathbf{\Delta}$):

$$\mathbf{P}_{[K]} = \left[\mathbf{p}_1, \ldots, \mathbf{p}_k, \ldots, \mathbf{p}_K\right] \tag{22}$$

$$\mathbf{Q}_{[K]} = \left[\mathbf{q}_1, \ldots, \mathbf{q}_k, \ldots, \mathbf{q}_K\right] \tag{23}$$

$$\mathbf{\Delta}_{[K]} = \operatorname{diag}\{\delta_1, \ldots, \delta_k, \ldots, \delta_K\} \ . \tag{24}$$

The matrix $\mathbf{A}$ reconstructed from the first K eigenvectors is denoted $\mathbf{A}_{[K]}$. It is obtained as:

$$\mathbf{A}_{[K]} = \mathbf{P}_{[K]}\mathbf{\Delta}_{[K]}\mathbf{Q}_{[K]}^T = \sum_k^K \delta_k \mathbf{p}_k \mathbf{q}_k^T \ , \tag{25}$$

(with δ_k being the k-th singular value).

The reconstructed matrix $\mathbf{A}_{[K]}$ is said to be optimal (in a least square sense) for matrices of rank K because it satisfies the following condition:

$$\left\|\mathbf{A} - \mathbf{A}_{[K]}\right\|^2 = \operatorname{trace}\left\{\left(\mathbf{A} - \mathbf{A}_{[K]}\right)\left(\mathbf{A} - \mathbf{A}_{[K]}\right)^\mathsf{T}\right\} = \min_{\mathbf{X}} \|\mathbf{A} - \mathbf{X}\|^2 \tag{26}$$

for the set of matrices $\mathbf{X}$ of rank smaller or equal to K. The quality of the reconstruction is given by the ratio of the first K eigenvalues (*i.e.,* the squared singular values) to the sum of all the eigenvalues. This quantity is interpreted as *the reconstructed proportion* or *the explained variance*, it corresponds to the inverse of the quantity minimized by Equation 27. The quality of reconstruction can also be interpreted as the squared coefficient of correlation (precisely as the R_v coefficient, see entry) between the original matrix and its approximation.

The GSVD minimizes an expression similar to Equation 27, namely

$$\mathbf{A}_{[K]} = \min_{\mathbf{X}} \left[\operatorname{trace}\left\{\mathbf{M}(\mathbf{A} - \mathbf{X})\mathbf{W}(\mathbf{A} - \mathbf{X})^\mathsf{T}\right\}\right] \ , \tag{27}$$

for the set of matrices $\mathbf{X}$ of rank smaller or equal to K.

4.1 SVD and General linear model

It can be shown that the SVD of a rectangular matrix gives the PCA of this matrix, with, for example, the factor scores being obtained as $\mathbf{F} = \mathbf{P}\boldsymbol{\Delta}$.

The adequate choice of matrices $\mathbf{M}$ and $\mathbf{W}$ makes the GSVD a very versatile tool which can implement the set of methods of linear multivariate analysis. For example, correspondence analysis (see entry) can be implemented by using a probability matrix (*i.e.,* made of positive or null numbers whose sum is equal to 1) along with two diagonal matrices $\mathbf{M} = \mathbf{D_r}$ and $\mathbf{W} = \mathbf{D_c}$ representing respectively the relative frequencies of the rows and the columns of the data matrix. The other multivariate techniques (*e.g.,* discriminant analysis, canonical correlation analysis, discriminant analysis) can be implemented with the proper choice of the matrices $\mathbf{M}$ and $\mathbf{W}$ (see, *e.g.,* Greenacre, 1984).

5 An example of singular value decomposition: Image compression

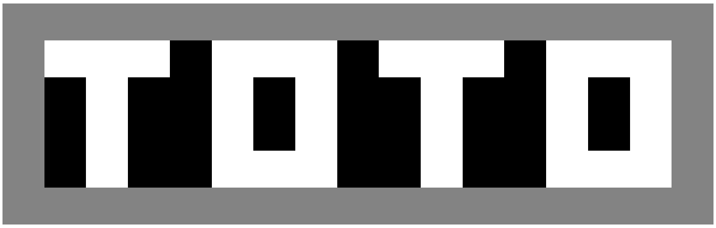

Figure 1: The matrix of Equation 28 displayed as a picture.

The SVD of a matrix is equivalent to PCA. We illustrate this property by showing how it can be used to perform image compression. Modern technology use digitized pictures, which are equivalent to a matrix giving the gray level value of each pixel. For example, the

Figure 2: A picture corresponding to a matrix in the order of $204 \times 290 = 59610$ pixels.

matrix:

$$\begin{bmatrix}1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&2&2&2&0&2&2&2&0&2&2&2&0&2&2&2&1\\1&0&2&0&0&2&0&2&0&0&2&0&0&2&0&2&1\\1&0&2&0&0&2&0&2&0&0&2&0&0&2&0&2&1\\1&0&2&0&0&2&2&2&0&0&2&0&0&2&2&2&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\end{bmatrix} \quad (28)$$

corresponds to the image in Figure 1.

In general, pictures coming from natural images have rather large dimensions. For example, the picture shown in Figure 2 corresponds to a matrix with 204 rows and 290 columns (therefore $204 \times 290 = 59610$ pixels). To avoid the problem of transmitting or storing the numerical values of such large images we want to represent the image with fewer numerical values than the original number of pixels.

Thus, one way of compressing an image is to compute the singular value decomposition and then to reconstruct the image by

Figure 3: The picture in Figure 2 built back with 25 pairs of singular vectors. (compression rate of $\approx 80\%$)

an approximation of smaller rank. This technique is illustrated in Figures 4 and 5, which show respectively the terms $\mathbf{p}_k\mathbf{q}_k^\mathsf{T}$ and the terms $\sum \mathbf{p}_k\mathbf{q}_k^\mathsf{T}$. As can be seen in Figure 4, the image is reconstructed almost perfectly (according to the human eye) by a rank 25 approximation. This gives a compression ratio of:

$$1 - \frac{25 \times (1 + 204 + 290)}{204 \times 290} = 1 - .2092 = .7908 \approx 80\% \,. \qquad (29)$$

References

[1] Abdi, H. (2003). Multivariate analysis. In M. Lewis-Beck, A. Bryman, & T. Futing (Eds): *Encyclopedia for research methods for the social sciences*. Thousand Oaks: Sage.

[2] Abdi, H., Valentin, D. (2006). *Mathématiques pour les sciences cognitives (Mathematics for cognitive sciences)*. Grenoble: PUG.

[3] Greenacre, M.J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

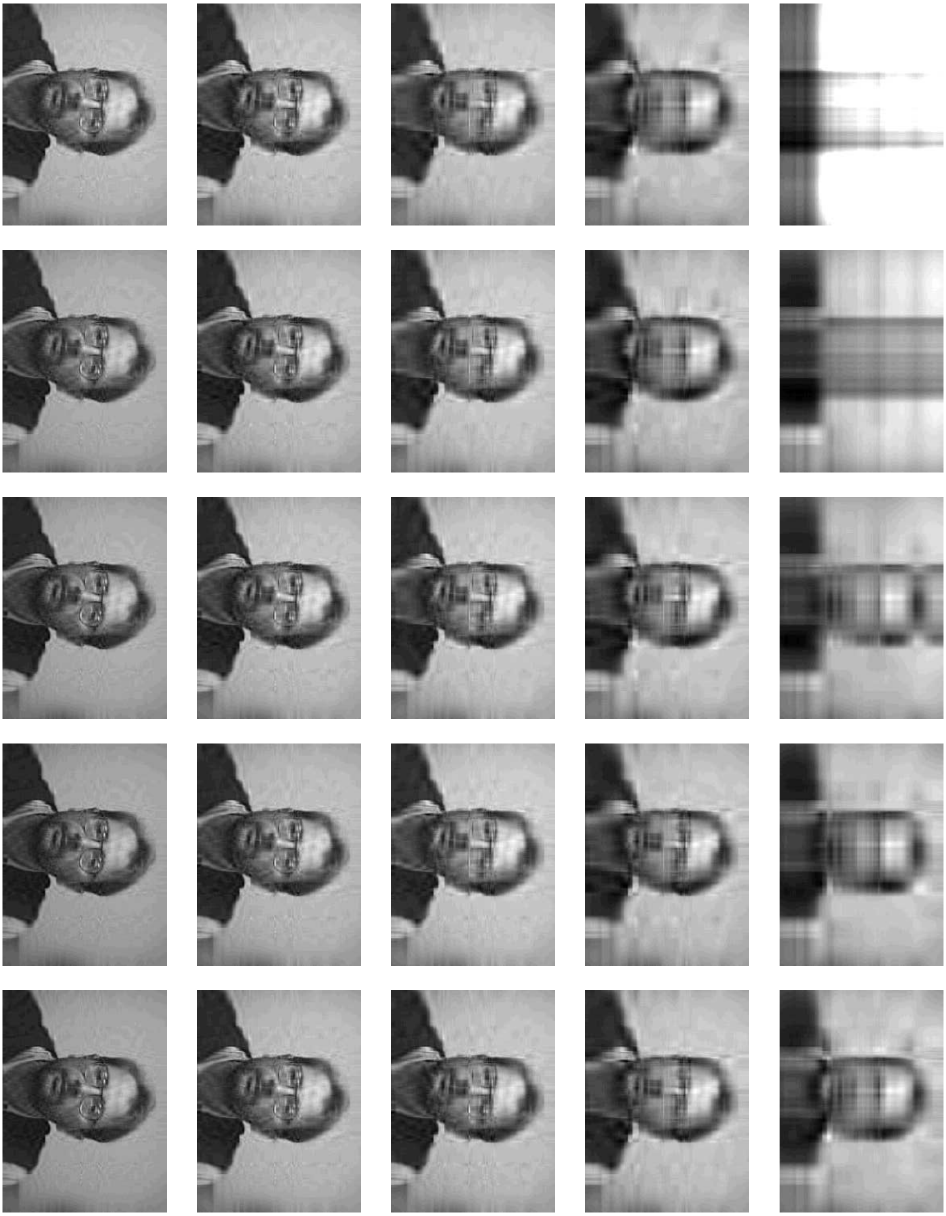

Figure 4: Reconstruction of the image in Figure 2. The percentages of explained variance are: 0.9347; 0.9512; 0.9641; 0.9748; 0.9792; 0.9824; 0.9846; 0.9866; 0.9881; 0.9896; 0.9905; 0.9913; 0.9920; 0.9926; 0.9931; 0.9936; 0.9940; 0.9944; 0.9947; 0.9950; 0.9953; 0.9956; 0.9958; 0.9961; 0.9963;

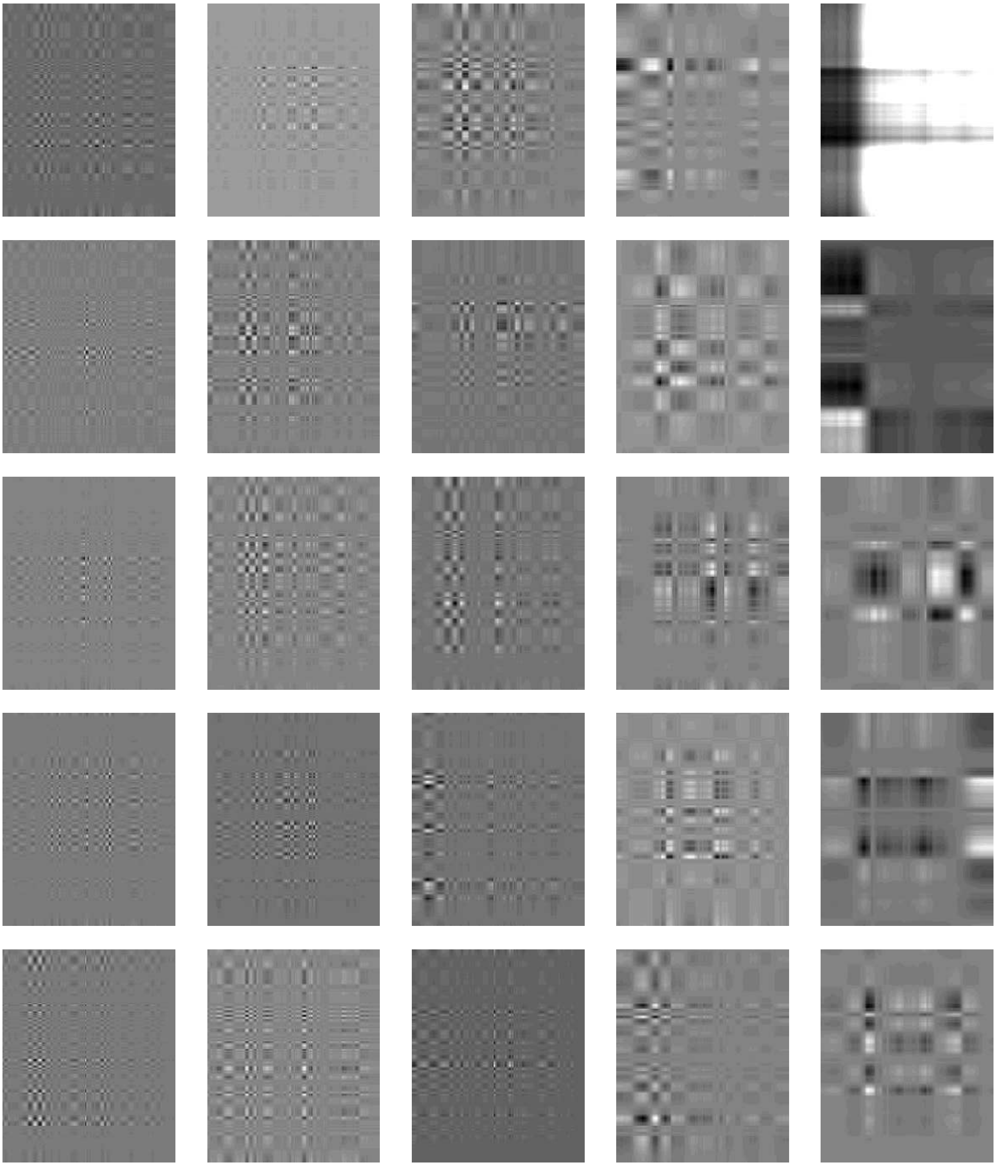

Figure 5: The terms $\mathbf{p}_k\mathbf{q}_k$ used to reconstruct the image in Figure 2 (see Figure 4). The eigenvalues (squared singular values) associated to each image are: 0.9347; 0.0166; 0.0129; 0.0107; 0.0044; 0.0032; 0.0022; 0.0020; 0.0015; 0.0014; 0.0010; 0.0008; 0.0007; 0.0006; 0.0005; 0.0005; 0.0004; 0.0004; 0.0003; 0.0003; 0.0003; 0.0003; 0.0002; 0.0002; 0.0002.

[4] Strang, G. (2003). *Introduction to linear algebra*. Cambridge (MA): Wellesley-Cambridge Press.