ON THE SINGULAR DECOMPOSITION OF MATRICES

Alina PETRESCU-NIŢĂ

Abstract

This paper is an original presentation of the algorithm of the singular decomposition (and implicitly, of the calculus of the pseudoinverse) of any matrix with real coefficients. In the same time, we give a geometric interpretation of that decomposition. In the last section, we present an application of the singular decomposition in a problem of Pattern Recognition.

1 Introduction

Let $A \in M_{m,n}(\mathbb{R})$ be a matrix; for any column vector $X \in \mathbb{R}^{\mathbf{n}} \simeq M_{\mathbf{n},1}(\mathbb{R})$, define $f_A : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{m}}, \ \ f_A(X) = AX$.

One may associate to A four remarkable vector spaces:

$$I(A) = Im(f_A), \ \ N(A) = Ker(f_A),$$

$$I(A^T) \ \text{ and } \ N(A^T) = \left\{ Y | A^T Y = 0 \right\}.$$

$I(A)$ is the vector subspace of $\mathbb{R}^{\mathbf{m}}$, generated by the columns of the matrix A, whereas $I(A^T)$ is the subspace of $\mathbb{R}^{\mathbf{n}}$, generated by the rows of A. If the rank of matrix A is $\rho(A) = r$, then $\dim I(A) = r, \ \dim N(A) = n - r, \ \dim I(A^T) = r, \ \dim N(A^T) = m - r$.

By convention, any vector $x \in \mathbb{R}^{\mathbf{n}}, \ x = (x_1, x_2, ..., x_{\mathbf{n}})$, can be identified with the corresponding column vector X; for any $x, y \in \mathbb{R}^{\mathbf{n}}$, the Euclidian scalar product is defined by $\langle x, y \rangle = X^T Y$.

The following properties are well-known:

1) The subspace $N(A)$ is orthogonal to $I(A^T)$ and $N(A^T)$ is orthogonal to $I(A)$.

2) The following orthogonal decompositions hold:

$$\mathbb{R}^{\mathbf{n}} = N(A) \oplus I(A^T), \ \mathbb{R}^{\mathbf{m}} = N(A^T) \oplus I(A).$$

3) The square matrix $A^T A$ of order n is symmetric and non-negatively defined (i.e. $X^T(A^T A)X \geq 0$, for any column vector X).

4) For any matrix $A \in M_{m,n}(\mathbb{R})$ of rank r, the number of positive eigenvectors of the matrix $A^T A$ is equal to r.

5) We suppose that $m \geq n$. Let $\lambda_1, \lambda_2, ..., \lambda_r$ be the positive eigenvalues of matrix $A^T A$ and $\sigma_k = \sqrt{\lambda_k}$, $1 \leq k \leq r$ singular numbers of the matrix A. These are called singular numbers of the matrix A. There exists an orthonormal basis $\{v_1, v_2, ..., v_{\mathbf{n}}\}$ of $\mathbb{R}^{\mathbf{n}}$ formed by the unitary eigenvectors of $A^T A$, such that $A^T A v_i = \sigma_i^2 v_i$, $1 \leq i \leq r$, and $A^T A v_j = 0$, $r+1 \leq j \leq n$. If we denote $u_i = \frac{1}{\sigma_i} A v_i$, $1 \leq i \leq r$, it follows that the vectors $v_1, v_2, ..., v_r$ form an orthonormal basis for $I(A^T)$ and $v_{r+1}, v_{r+2}, ..., v_{\mathbf{n}}$ form an orthonormal basis for $N(A)$. Further, $u_1, u_2, ..., u_r$ form an orthonormal basis for $I(A)$, which can be extended to an orthonormal basis $u_1, u_2, ..., u_r, u_{r+1}, u_{r+2}, ..., u_{\mathbf{m}}$ of $\mathbb{R}^{\mathbf{m}}$; the added vectors $u_{r+1}, u_{r+2}, ..., u_m$ form an orthonormal basis for $N(A^T)$.

6) The orthogonal matrices $U = (u_1|u_2|...|u_{\mathbf{m}})$, $V = (v_1|v_2|...|v_{\mathbf{n}})$ are invertible (the inverse is just the transposed matrix). We have:

$$AV = (Av_1|Av_2|...|Av_{\mathbf{n}}) = (\sigma_1 u_1|\sigma_{\mathbf{2}} u_2|...|\sigma_r u_r|0|...|0),$$

$$\sigma_{\mathbf{i}} \neq 0, i = 1, 2, ..., r$$

and if we denote

$S = diag(\sigma_1, \sigma_2, ..., \sigma_r)$ and $\Sigma = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}$ being an a $m \times n$ - matrix, it follows that $AV = U\Sigma$ and following relation holds:

$$A = U\Sigma V^T. \tag{1}$$

This is the **singular decomposition** of matrix A.

If $A = 0$, then $\Sigma = 0$.

7) Until now we have supposed that $m \geq n$. If $A \in M_{m,n}(\mathbb{R})$ and $m < n$, then we may consider the $n \times m$ - matrix $B = A^T$. According to the properties 5) and 6), we have the singular decomposition $B = U_1 \Sigma_1 V_1^T$ and it follows

$$A = B^T = V_1 \Sigma_1 U_1^T. \tag{1'}$$

The main application of the singular decomposition is the explicit way to compute of the pseudoinverse $A^\dagger$ of an arbitrary non-null matrix $A \in M_{m,n}(\mathbb{R})$, namely, if $A = U\Sigma V^T$ (according to (1)), then $A^\dagger = V\Sigma^\dagger U^T$, where

$$\Sigma^\dagger = \begin{pmatrix} S^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in M_{n,m}(\mathbb{R}) \quad \text{and} \quad S^{-1} = diag\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, ..., \frac{1}{\sigma_r}\right)$$

NOTE. One knows that for any $A \in M_{m,n}(R)$, there is $A^+ \in M_{n,m}(\mathbb{R})$, called the pseudoinverse of A, such that for any $y \in \mathbb{R}^m$, the minimum of the euclidian norm $||Ax - y||$ is attained iff $x \in A^+y$.

2 The Singular Decomposition Algorithm

One knows that any rectangular matrix (or a square matrix, invertible or not) admits a singular decomposition given by (1) or (1').

Suppose now that the matrix $A \in M_{m,n}(\mathbb{R})$ is given, and $m \geq n$. If $m < n$, then the algorithm applies to the matrix A^T.

Step 1. Compute the symmetric matrix $A^T A \in M_n(\mathbb{R})$ and determine the nonzero eigenvalues $\lambda_1, \lambda_2, ..., \lambda_r$, as well as the singular numbers $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, ..., \sigma_r = \sqrt{\lambda_r}$, where $\rho(A) = r$.

Step 2. Determine an orthonormal basis $\{v_1, v_2, ..., v_\mathbf{n}\}$ of $\mathbb{R}^\mathbf{n}$, formed by the unitary eigenvectors of $A^T A$ and denote by $V \in M_n(\mathbb{R})$ the orthogonal matrix whose columns are formed by the vectors $v_1, v_2, ..., v_\mathbf{n}$.

Step 3. Compute the column unitary vectors $u_i = \frac{1}{\sigma_i} A v_i$ for $1 \leq i \leq r$ and complete them to an orthonormal basis $\{u_1, u_2, ..., u_r, u_{r+1}, u_{r+2}, ..., u_\mathbf{m}\}$ of $\mathbb{R}^\mathbf{m}$. Denote by $U \in M_m(\mathbb{R})$ the orthogonal matrix formed by the column vectors

$$u_1, u_2, ..., u_r, u_{r+1}, u_{r+2}, ..., u_\mathbf{m}.$$

Step 4. Taking $S = diag(\sigma_1, \sigma_2, ..., \sigma_r)$ and defining the $m \times n$ - matrix Σ having S in the left upper corner, that is

$$\Sigma = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix},$$

one obtains the singular decomposition (1).

Example 1 Take $A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$ hence $m = 2$, $n = 2$, $r = 1$.

The matrix $A^T A$ has the eigenvalues $\lambda_1 = 10, \lambda_2 = 0$, with the unitary eigenversors $v_1 = \frac{1}{\sqrt{5}}(1,2)^T$; $v_2 = \frac{1}{\sqrt{5}}(-2,1)^T$. Then $u_1 = \frac{1}{\sqrt{10}}Av_1 = \frac{1}{\sqrt{2}}(1,1)$ and take $u_2 = \frac{1}{\sqrt{2}}(1,-1)$. So,

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, V = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}, \Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix}$$

and finally, $A = U\Sigma V^T$.

Example 2 For $B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ and $A = B^T$, $A^T A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$; $\lambda_1 = 3, \lambda_2 = 1$ and the singular numbers of A are $\sigma_1 = \sqrt{3}, \sigma_2 = 1$ $(r = 2)$. Hence, the unitary eigenvectors for $A^T A$ are:

$$v_1 = \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}} \right)^T \text{ and } v_2 = \left(\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}} \right)^T. \text{ Take } u_1 = \tfrac{1}{\sqrt{3}} A v_1 =$$

$\left(\tfrac{1}{\sqrt{6}}, \tfrac{2}{\sqrt{6}}, \tfrac{1}{\sqrt{6}} \right)^T, u_2 = A v_2 = \left(\tfrac{1}{\sqrt{2}}, 0, -\tfrac{1}{\sqrt{2}} \right)^T$ and we complete u_1, u_2 to an orthonormal basis of $\mathbb{R}^3$. We take $u_3 = (a, b, c)^T$ with unknown components and impose the condition $u_3 \perp u_1, u_3 \perp u_2$ and $a^2 + b^2 + c^2 = 1$. It follows $u_3 = \left(\tfrac{1}{\sqrt{3}}, -\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}} \right)^T$ and denote:

$$U = \begin{pmatrix} \dfrac{1}{\sqrt{6}} & \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{3}} \\[2mm] \dfrac{2}{\sqrt{6}} & 0 & -\dfrac{1}{\sqrt{3}} \\[2mm] \dfrac{1}{\sqrt{6}} & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{3}} \end{pmatrix},$$

$$V = \begin{pmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\[2mm] \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

finally, the singular decomposition $A = U \Sigma V^T$ and $B = V \sum^T U^T$.

3   Geometric Interpretation of Singular Decomposition

From geometrical point of view, any orthogonal matrix $U \in M_m(\mathbb{R})$ corresponds to a rotation of space $\mathbb{R}^{\mathbf{m}}$. For $m = 2$, an orthogonal matrix $U \in M_2(\mathbb{R})$ has the form

$$U = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \theta \in \mathbb{R}$$

and the application $f_U : \mathbb{R}^2 \to \mathbb{R}^2, f_U(x, y) = (x', y')$ becomes $x' = x \cos \theta - y \sin \theta, y' = x \sin \theta + y \cos \theta$. This is the plane rotation formulae with the angle θ, around centered in the origin. This fact is generalized for upper dimensions.

Any matrix of type S (or Σ) corresponds, from a geometrical point of view, to a scale change. For instance, if $n = 2$ and $S = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$, the application $f_S : \mathbb{R}^2 \to \mathbb{R}^2, f_S(x, y) = (x', y')$, becomes $x' = \sigma_1 x$ and $y' = \sigma_2 y$, and we get the plane scale change formulae.

Proposition *Any linear application $f : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{n}}$ is a composition of a rotation with a scale change, followed by another rotation.*

Proof. Let A be the associated matrix of the linear application f with respect to the canonical basis. According to **1**, the singular decomposition of the matrix A has the form $A = U\Sigma V^T$, with U and V orthogonal matrices and Σ a diagonal matrix.

Then

$$f = f_A = f_U \circ f_\Sigma \circ f_V. \tag{2}$$

The maps f_U and f_V correspond to orthogonal matrices and they represent rotations, whereas f_Σ is a scale change, namely

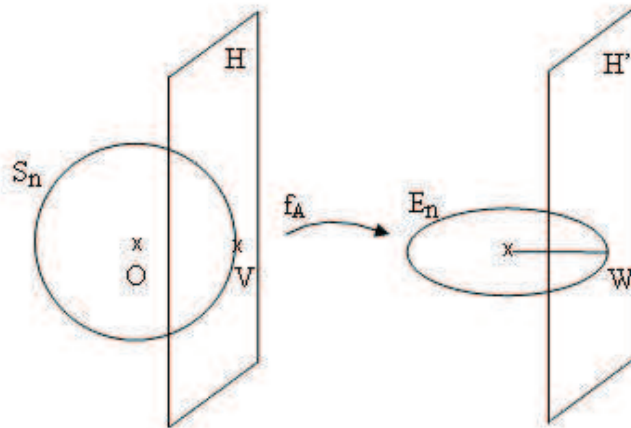$$f_\Sigma(x_1, x_2, ..., x_n) = (\sigma_1 x_1, ..., \sigma_r x_r, 0, ..., 0).$$

Figure 1. The image of the unit sphere S_n on f_A

Relation (2) correspond to the statement of the proposition.

Let $A \in M_n(\mathbb{R})$ be a nonsingular square matrix (hence $m = n = r$) and $f_A : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{n}}$ be a linear application associated to A in the canonical basis of $\mathbb{R}^{\mathbf{n}}$. Through the application f_A, the unit sphere $S_{\mathbf{n}} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is transformed into an n-dimensional ellipsoid $E_n = f_A(S_n)$.

Indeed, if $y = f_A(x)$, it follows that $Y = AX$, $X = A^{-1}Y$ hence

$$1 = \|X\|^2 = \|A^{-1}Y\|^2 = \langle A^{-1}Y, A^{-1}Y \rangle = Y^T (A^{-1})^T A^{-1} Y,$$

and the matrix $C = (A^{-1})^T A^{-1}$ is positively by defined, hence the set $E_n = \{Y = AX/X \in S_n\}$ defines an ellipsoid.

The recursive construction of the orthonormal bases $v_1, v_2, ..., v_n$ and $u_1, ..., u_n$ has also a geometric interpretation, which is presented in the sequel. Let w be

a radial vector of maximal length in the ellipsoid and $v = A^{-1}w$. If we denote by H the tangent hyperplane at v to the unitary sphere S_n and $H' = f_A(H)$, then it follows that the hyperplane H' is tangent at w to the ellipsoid E_n (see the figure 1).

Indeed, we have $w \in H'$ and H' has just one common point with the ellipsoid E_n (otherwise, since the application f_A is bijective, it would follow that H is not tangent to the sphere).

Moreover, $w \perp H'$. We take $v_1 = v, u_1 = \frac{w}{\|w\|}$. Considering the restriction g of the linear application f_A to H, we obtain a linear application $g : H \to H'$, for which we can do the previous construction. This geometric interpretation leads to the singular decomposition (1) without appealing the study of matrix $A^T A$.

4   An application of the Singular Decomposition to the Classification of 2D Images

Let $A \in M_{m,n}(\mathbb{R})$ be the gray levels matrix of a 2D black-white image (for example: a photography, a map, a black-white TV image, etc.). Such a matrix can be obtained by splitting the image with a rectangular network, and associate to each node (i,j), $1 \le i \le m, 1 \le j \le n$, the gray level expressed as an integer number in the range between 0 and 63, where, for example, 0 stands for "absolute white" and 63 stands for "absolute black".

Let us consider the singular decomposition

$$A = U\Sigma V^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T, \ \ r = \rho(A),$$

where $\lambda_1 > \lambda_2 > ... > \lambda_r > 0$ and $\lambda_1^2, \lambda_2^2, ..., \lambda_r^2$ are the nonzero eigenvalues of matrix AA^T. If $A = (a_{ij})$, $1 \le i \le m, 1 \le j \le n$, then the Frobenius norm $\|A\|_F = \left(\sum_{i,j} a_{ij}^2 \right)^{1/2}$ can be called the *energy* of the considered image.

If the "small" eigenvalues are eliminated, we obtain and approximation

$$A' = \sum_{i=1}^{k} \lambda_i u_i v_i^T, (k << r) \ \ \text{and} \ \ \|A - A'\|_F = \left(\sum_{i=k+1}^{r} \lambda_i^2 \right)^{1/2}.$$

If $B \in M_{m,n}(\mathbb{R})$ is another matrix, then matrix $\check{B} = U \check{\Sigma} V^T = \sum_{i=1}^{r} \bar{\lambda}_i u_i v_i^T$ can be called the projective image of B on A, $\overline{B} = U.\overline{\Sigma}.V^T = \sum_{i=1}^{r} \overline{\lambda}_i u_i v_i^T$, where $\overline{\Sigma} = \left(\overline{\lambda_1}, ..., \overline{\lambda_r}, 0, ..., 0 \right)$ and $\overline{\lambda}_i = u_i^T B v_i, 1 \le i \le \rho$. We have:

$$\left\| B - \overline{B} \right\|_F \le \|A - B\|_F + \left(\sum_{i=1}^{r} \left(\lambda_i - \overline{\lambda_i} \right)^2 \right)^{1/2}.$$

For similar 2D images of the same class, the distance between the associated matrices (i.e. the Frobenius norm of difference of the matrices) is also

"small"; passing to the projective images, these are smaller because $\left\|\overline{B} - \overline{C}\right\|_F \leq \left\|B - C\right\|_F$ for any two matrices $B, C \in M_{m,n}(\mathbb{R})$.

Let us suppose that, for an image class ω we have the learning sample $A_1, A_2, ..., A_N \in M_{m,n}(\mathbb{R})$. The average is given by $\mu = \frac{1}{N}(A_1 + ... + A_N)$ and like above, the singular decomposition of average is $\mu = U\Sigma V^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T$. Similarly we get the projective images $\overline{A}_1, ..., \overline{A}_N$ on μ, hence $\overline{A}_i = \sum_{j=1}^{r} x_j^{(i)} u_j v_j^T$, $1 \leq i \leq N$, where $x_j^{(i)} = u_j^T A_i v_j, 1 \leq j \leq r$.

The vector $X_i = \left(x_1^{(i)}, ..., x_r^{(i)}\right)$ can be interpreted as the coordinates vector of the projective image on μ of matrices A_i, $1 \leq i \leq N$.

The algorithm of supervised classification of images

Let $\omega_1, \omega_2, ..., \omega_M$ be M classes of images (already existing classes). We suppose that each class w_i is represented by N_i matrices of learning samples $A_1^{(i)}, A_2^{(i)}, ..., A_{N_i}^{(i)}$ belonging to $M_{m,n}(\mathbb{R})$.

Step1. Compute the average $\mu_i = \frac{1}{N_i}\left(A_1^{(i)} + ... + A_{N_i}^{(i)}\right)$ and the singular decomposition of this matrix, hence the set of matrices

$$u_j^{(i)}, v_j^{(i)}, 1 \leq j \leq k, \quad k \leq \min(m, n), \quad 1 \leq i \leq M.$$

Step2. Compute the vectors of the coordinates of the projective image on μ_i, i.e. $X_j^{(i)} = \left(x_{j_1}^{(i)}, ..., x_{j_r}^{(i)}\right)$, where:

$$x_{j_p}^{(i)} = \left(u_p^{(i)}\right)^T A_j^{(i)} v_p^{(i)}, 1 \leq p \leq k, 1 \leq j \leq N_i, 1 \leq i \leq M.$$

Step3. Compute the "center" of the classes ω_i by $X_C^{(i)} = \frac{1}{N_i}\sum_{j=1}^{N_i} X_j^{(i)}$.

Step4. (the recognition step) For any unclassified new 2D image $F \in M_{m,n}(\mathbb{R})$, compute the projective images of F on $\mu_i, 1 \leq i \leq M$, and the corresponding coordinates vectors $Y_1, ..., Y_M$. If we denote $z_p^{(i)} = \left(u_p^{(i)}\right)^T F v_p^{(i)}, 1 \leq p \leq k$, we have $Y_i = \left(z_1^{(i)}, ..., z_k^{(i)}\right), 1 \leq i \leq M$.

If $\min_{1 \leq i \leq M}\left\|Y_i - X_C^{(i)}\right\|_F$ is reached for an index $i = i_0$, (not necessary unique) then image F is places in class ω_{i_0}.

References

[1] Ben-Israel A., Greville T.N.E., *Generalized Inverses; Theory and Applications*, Springer Verlag N.Y., Inc., 2003.

[2] Niţă, A., *Generalized inverse of a matrix with applications to optimization of some systems (Ph. D. Thesis, in Romanian)*, University of Bucharest, Faculty of Mathematics and Computer Science, 2004.

[3] Strang, G., *Linear algebra and applications*, Academic Press, 1976.

[4] Wang, J., *A recurrent neural network for real-time*, Appl.Math. Computer, **55**(1993), 89-100.

[5] Wang, J., *Recurrent neural networks for computing pseudoinverses of rank-deficient matrices*, Appl.Math. Computer, **18**(1997) 1479-1493.

University POLITEHNICA of Bucharest
Department of Mathematics
Splaiul Independenţei 313, Ro - Bucharest, Romania
e-mail: anita@euler.math.pub.ro