Singular Value Decomposition

J. H. Caltenco, J. López-Bonilla, B. E. Carvajal-Gámez, P. Lam-Estrada,

1 ESIME-Zacatenco, Instituto Politécnico Nacional (IPN), Edif. 5, CP 07738, México DF; jlopezb@ipn.mx
2 SEPI-ESCOM, IPN, Av. Bátiz S/N, Col. Nueva Industrial Vallejo, CP 07738, México DF,
3 ESFM-IPN, Depto. de Matemáticas, Edif. 9, Col. Lindavista, CP 07738, México DF

Keywords: Pseudoinverse of a matrix, SVD, Compatibility of linear systems

Abstract. We study the SVD of an arbitrary matrix $A_{n,m}$, especially its subspaces of activation, which leads in natural manner to pseudoinverse of Moore-Bjenhammar-Penrose. Besides, we analyze the compatibility of linear systems and the uniqueness of the corresponding solution, and our approach gives the Lanczos classification for these systems.

1. Introduction

For any real matrix $A_{n,m}$, Lanczos [1] constructs the matrix:

$$S_{(n+m)\times(n+m)} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix},$$

and he studies the eigenvalue problem:

$$S\tilde{\omega} = \lambda\tilde{\omega},$$

where the proper values are real because $S$ is a real symmetric matrix. Besides:

$$\text{rank } A = p = \text{Number of positive eigenvalues of } S,$$

such that $1 \leq p \leq \min(n,m)$. Then the singular values or canonical multipliers, thus called by Picard [2] and Sylvester [3], respectively, follow the scheme:

$$\lambda_1, \lambda_2, \ldots, \lambda_p, -\lambda_1, -\lambda_2, \ldots, -\lambda_p, 0, 0, \ldots, 0,$$

that is, $\lambda = 0$ has the multiplicity $n + m - 2p$. Only in the case $p = n = m$ can occur the absence of the null eigenvalue.

The proper vectors of $S$, named „essential axes“ by Lanczos, can be written in the form:

$$\tilde{\omega}_{(n+m)\times 1} = \begin{pmatrix} \overline{v} \\ \overline{v} \end{pmatrix}^n_m,$$

then (1) and (2) imply the Modified Eigenvalue Problem:

$$A_{n,m}\tilde{v}_{mx1} = \lambda\tilde{u}_{mx1}, \quad A_{m,n}^T\tilde{u}_{mx1} = \lambda\tilde{v}_{mx1},$$

Thus

$$A^TA\tilde{v} = \lambda^2\tilde{v}, \quad AA^T\tilde{u} = \lambda^2\tilde{u},$$

with special interest in the associated vectors with the positive eigenvalues because they permit to introduce the matrices:
This relation tells that in the construction of A we do not need information about the null proper value; the information from \( \omega \) is important to study the existence and uniqueness of the solutions for a linear system associated to A. This approach of Lanczos is similar [7] to Schmidt [8] and Jordan [9, 10] methods; we can consider that Jordan, Sylvester [3] and Beltrami [11] are the founders of the SVD [12], and there is abundant literature [13-25] on this matrix factorization and its applications.

In Sec. 2, we realize an analysis of the proper vectors \( \overrightarrow{\omega}_j, j = 1, ..., n + m \), associated to the eigenvalues (4), which leads to the subspaces of activation of A with the pseudoinverse of Moore [26]-Bjerhammar [27]-Penrose [28]. In Sec. 3, we study the compatibility of linear systems, with special emphasis in the important participation of the null singular value and its corresponding eigenvectors.

2. Subspaces of activation and natural inverse matrix

From (6), the proper vectors associated with the positive eigenvalues verify:

\[
A \overleftarrow{\omega}_j = \lambda_j \overleftarrow{\omega}_j, \quad A^T \overleftarrow{\omega}_j = \lambda_j \overleftarrow{\omega}_j, \quad j = 1, ..., p
\]

then

\[
A(\overrightarrow{\omega}_j) = (-\lambda_j) \overrightarrow{\omega}_j, \quad A^T \overrightarrow{\omega}_j = (-\lambda_j)(-\overrightarrow{\omega}_j),
\]

that is

\[
S \left( \frac{\overrightarrow{\omega}_k}{\overrightarrow{\omega}_k} \right) = \lambda_k \left( \frac{\overrightarrow{\omega}_k}{\overrightarrow{\omega}_k} \right) \quad \text{implies} \quad S \left( \frac{-\overleftarrow{\omega}_k}{-\overleftarrow{\omega}_k} \right) = (-\lambda_k) \left( \frac{-\overleftarrow{\omega}_k}{-\overleftarrow{\omega}_k} \right).
\]

Therefore, the eigenvectors \( \frac{\overrightarrow{\omega}_j}{\overrightarrow{\omega}_j} \) and \( \frac{-\overleftarrow{\omega}_j}{-\overleftarrow{\omega}_j} \) correspond to the proper values \( \lambda_1, ..., \lambda_p \) and \(-\lambda_1, ..., -\lambda_p\), respectively. Thus we must have \( n + m - 2p \) eigenvectors connected to \( \lambda = 0 \), which we denote by \( \overrightarrow{\omega}^{(0)} \), and from (6):

\[
\overrightarrow{\omega}^{(0)}_j = \begin{pmatrix} u^{(0)}_j \\ v^{(0)}_j \end{pmatrix} \quad \text{and} \quad A^T \overrightarrow{\omega}^{(0)}_j = 0, \quad j = 1, ..., n - p,
\]

\[
\overrightarrow{\omega}^{(0)}_{(n-p)+k} = \begin{pmatrix} u^{(0)}_{n-k} \\ v^{(0)}_{n-k} \end{pmatrix} \quad \text{and} \quad A \overrightarrow{\omega}^{(0)}_k = 0, \quad k = 1, ..., m - p.
\]
The conditions (14) and (15) can be multiplied by $A$ and $A^T$, then $\tilde{u}_j^{(0)}$ and $\tilde{v}_k^{(0)}$ are eigenvectors of the Gram matrices $AA^T$ and $A^TA$:

\[
(AA^T)_{n \times n} \tilde{u}_j^{(0)} = \delta_{jj}, \quad (A^TA)_{m \times m} \tilde{v}_k^{(0)} = \delta_{kk},
\]

(16)

but by (7) these matrices have $p$ proper vectors for $\lambda_1, \ldots, \lambda_p$, therefore only there are $n - p$ and $m - p$ vectors $\tilde{u}_j^{(0)}$ and $\tilde{v}_k^{(0)}$ that can be selected with orthonormality:

\[
\tilde{u}_j^{(0)} \cdot \tilde{u}_r^{(0)} = \delta_{jr}, \quad \tilde{v}_k^{(0)} \cdot \tilde{v}_q^{(0)} = \delta_{kq},
\]

(17)

that is, $\tilde{u}_j^{(0)} \cdot \tilde{u}_k^{(0)} = \delta_{jk}$, then \( \{\tilde{u}_j^{(0)}\} \) and \( \{\tilde{v}_k^{(0)}\} \) are bases for the Kernel $A^T$ and Kernel $A$, respectively.

If into (14) we employ (10) [SVD of $A$] results $V\Lambda U^T \tilde{u}_j^{(0)} = \tilde{0}$, whose multiplication by the left with $\Lambda^{-1}V^T$ [remembering that $UU^T \tilde{u}_j^{(0)} = \tilde{0}$] gives the compatibility condition:

\[
V^T \tilde{u}_j^{(0)} = \tilde{0} \quad \Rightarrow \quad \tilde{u}_r \cdot \tilde{u}_j^{(0)} = 0, \quad r = 1, \ldots, p; \quad j = 1, \ldots, n - p,
\]

(18)

Equivalently

\[
\mathrm{Col} \ U \perp \tilde{u}_k^{(0)}, \quad k = 1, \ldots, n - p.
\]

(19)

Similarly, if we use SVD into (15) and we multiply by $\Lambda^{-1}U^T$:

\[
V^T \tilde{v}_k^{(0)} = \tilde{0}, \quad \tilde{v}_r \cdot \tilde{v}_k^{(0)} = 0, \quad r = 1, \ldots, p; \quad k = 1, \ldots, m - p
\]

(20)

\[
\therefore \quad \mathrm{Col} \ V \perp \tilde{v}_j^{(0)}, \quad j = 1, \ldots, m - p.
\]

(21)

It is convenient to make two remarks:

a) From $A = U\Lambda V^T$ is evident that the matrices $U$, $\Lambda$, and $V$ permit to construct $A$, but is useful to know more about the structure of $A$ and its transpose:

\[
A = (\tilde{a}_1 \ldots \tilde{a}_m), \quad A^T = (\tilde{c}_1 \ldots \tilde{c}_n),
\]

(22)

where $(\tilde{a}_j)_{m \times 1}$ and $(\tilde{c}_k)_{m \times 1}$ are the corresponding columns. Then from (10) we obtain the expressions:

\[
\tilde{a}_j = \lambda_1 \tilde{u}_1^{(j)} u_1 + \cdots + \lambda_p \tilde{u}_p^{(j)} u_p, \quad j = 1, \ldots, m, \quad \tilde{c}_k = \lambda_1 \tilde{v}_1^{(k)} v_1 + \cdots + \lambda_p \tilde{v}_p^{(k)} v_p, \quad k = 1, \ldots, n
\]

(23)

with the notation:

\[
u_r^{(j)} = j \text{ th - component of } \tilde{v}_r,
\]

(24)

and similar for $u_r^{(k)}$; we observe that $\tilde{c}_k^T$ are the rows of $A$. 


and similar for \( u^{(k)}_r \); we observe that \( c^T_k \) are the rows of \( A \).

From (23) are immediate the equalities of subspaces:

\[
\text{Col } A = \text{Col } U, \quad \text{Row } A = \text{Col } V, \quad (25)
\]

but \( \dim \text{Col } U = \dim \text{Col } V = p \), then:

\[
\text{rank } A = \dim \text{Col } A = \dim \text{Row } A = p \quad (26)
\]

in accord with (3).

\(-\) We have the rank-nullity theorem [29-31]:

\[
\begin{align*}
\dim (\text{Kernel } A) &+ \text{rank } A = m, \quad (27) \\
\dim (\text{Kernel } A^T) &+ \text{rank } A^T = n, \quad (28)
\end{align*}
\]

but \( \text{rank } A^T = \text{rank } A = p \), then \( \dim (\text{Kernel } A^T) = n - p \) in harmony with the \((n - p)\) vectors \( u_j^{(0)} \) verifying (14).

If \( A_{\text{norm}} \) acts on an arbitrary vector \( \vec{x} \in E_m \) produces a vector \( \vec{y} \in E_n \), with the decompositions:

\[
\begin{align*}
\vec{x} &= x^{(0)} + x_{CV}, \\
\vec{y} &= y^{(0)} + y_{CU}, \\
\end{align*}
\]

\[
\begin{align*}
\vec{x}^{(0)} &\in \text{Kernel } A, \quad \vec{x}_{CV} \in \text{Col } V, \quad A\vec{x}^{(0)} = 0, \quad \vec{x}^{(0)} \cdot \vec{x}_{CV} = 0, \quad (29)
\]

where

\[
\begin{align*}
\vec{y}^{(0)} &\in \text{Kernel } A^T, \quad \vec{y}_{CU} \in \text{Col } U, \quad A^T\vec{y}^{(0)} = 0, \quad \vec{y}^{(0)} \cdot \vec{y}_{CU} = 0, \quad (30)
\end{align*}
\]

Therefore, \( A\vec{x} = A\vec{x}_{CV} = \vec{y} \) and in the construction of \( \vec{y} \) we lost the information about \( \vec{x}^{(0)} \), then it is not possible to recover \( \vec{x} \) from \( \vec{y} \); that is, it is utopian to search for an “inverse matrix” acting on \( \vec{y} \) to give \( \vec{x} \). However, when \( \vec{x}^{(0)} = \vec{0} \) and \( \vec{y}^{(0)} = \vec{0} \) we can introduce a “natural inverse matrix”, thus named it by Lanczos, which coincides with the pseudoinverse of Moore [16]-Bjerhammar [27]-Penrose [28]:

“Any matrix \( A_{\text{norm}} \), restricted to its subspaces of activation, always can be inverted”.

(31)

In fact, if \( \vec{x} \in \text{Col } V \) is an arbitrary vector, \( \vec{x} = q_1\vec{v}_1 + \cdots + q_p\vec{v}_p \), then from (6):

\[
A\vec{x} = \lambda_1 q_1\vec{u}_1 + \cdots + \lambda_p q_p\vec{u}_p = \vec{y} \in \text{Col } U, \quad (32)
\]

and now we search the inverse natural \( A_{\text{norm}}^{-1} \) such that:

\[
A_{\text{norm}}^{-1} \vec{y} = \vec{x}, \quad (33)
\]
or more general:

\[ A_N^{-1} A \hat{x} = \hat{x}, \quad \forall \; \hat{x} \in \text{Col} \; V, \quad A A_N^{-1} \hat{y} = \hat{y}, \quad \forall \; \hat{y} \in \text{Col} \; U. \]  

(34)

If the decomposition (10) is applied to (32) we deduce the natural inverse matrix:

\[ A_{N \times n}^{-1} = V_{m \times m} A_{p \times p}^{-1} U_{p \times n}^T. \]  

(35)

satisfying (33) and (34). With (35) is easy to prove the properties [24, 32]:

\[ A A_N^{-1} A = A, \quad A_N^{-1} A A_N^{-1} = A_N^{-1}, \quad (A A_N^{-1})^T = A A_N^{-1}, \quad (A_N^{-1} A)^T = A_N^{-1} A, \]  

(36)

which characterize the pseudoinverse of Moore-Bjerhammar-Penrose, that is, the inverse matrix [32, 33] of these authors coincides with the natural inverse (35) deduced by Lanczos [1, 4-6], which characterize the pseudoinverse of Moore-Bjerhammar-Penrose, that is, the inverse matrix [32, 33] of these authors coincides with the natural inverse (35) deduced by Lanczos [1, 4-6].

In the SVD only participate the positive proper values of \( \Sigma \), without the explicit presence of the vectors \( \vec{u}^{(0)}_j \) and \( \vec{v}^{(0)}_k \) associated with the null eigenvalue, then it is natural to investigate the role performed by the information related with \( \lambda = 0 \). In Sec. 3, we study linear systems where \( A \) is the corresponding matrix of coefficients, and we exhibit that the \( \vec{u}^{(0)}_j \) permit to analyze the compatibility of such systems; besides, when they are compatible, then with the \( \vec{v}^{(0)}_k \) we search if the solution is unique. In other words, the null eigenvalue does not participates when we consider to \( A \) as an algebraic operator and we construct its factorization (10), but \( \lambda = 0 \) is important if \( A \) acts as the matrix of coefficients of a linear system.

3. Compatibility of linear systems

A linear system of \( n \) equations with \( m \) unknowns can be written in the matrix form:

\[ A_{n \times m} \hat{x}_{n \times 1} = \vec{b}_{n \times 1} \]  

(37)

where (10) implies that \( U \Lambda V^T \hat{x} = \vec{b} \), whose multiplication by \( \vec{u}^{(0)}_j \) gives the compatibility conditions:

\[ \vec{u}^{(0)}_j \cdot \vec{b} = 0, \quad j = 1, \ldots, n - p \]  

(38)

due to (19). Then the system (37) is compatible if \( \vec{b} \) is orthogonal to all independent solutions of the adjoint system \( A^T \vec{u} = \vec{0} \), therefore:

\[ "A \hat{x} = \vec{b} \] has solution if \( \vec{b} \in \text{Col} \; U", \]  

(39)

which is the traditional formulation [6] of the compatibility condition for a linear system given. From (25) and (39) is clear that \( A \) and the augmented matrix \( (A \; \vec{b}) \) have the same column space:

\[ \text{Col} \; A = \text{Col} \; (A \; \vec{b}) = \text{Col} \; U, \]  

(40)

thus at the books [32] we find the result:

\[ "A \hat{x} = \vec{b} \] is compatible if \( \text{rank} \; A = \text{rank} \; (A \; \vec{b})". \]  

(41)
If \( \vec{b} \in \text{Col} \ U \), then from (11):

\[
\vec{b} = b^{(1)} \vec{u}_1 + \cdots + b^{(p)} \vec{u}_p = A \vec{Q}, \quad \vec{Q} = \frac{b^{(1)}}{\lambda_1} \vec{v}_1 + \cdots + \frac{b^{(p)}}{\lambda_p} \vec{v}_p,
\]

and (37) leads to:

\[
A(\vec{x} - \vec{Q}) = \vec{0}.
\]

The set of solutions of (43) is the Kernel \( A \) with dimension \((m - p)\) due to (27), therefore (43) has the unique solution \( \vec{x} - \vec{Q} = \vec{0} \) when \( p = m \), that is, when rank \( A \) coincides with the number of unknowns we have not vectors \( \vec{v}_k^{(0)} \neq \vec{0} \) verifying \( A \vec{v}_k^{(0)} = \vec{0} \). Then:

"The compatible system \( A \vec{x} = \vec{b} \) has unique solution only when \( p = m \),

\[
x_r = Q^{(r)} = \frac{b^{(1)}}{\lambda_1} v_1^{(r)} + \cdots + \frac{b^{(p)}}{\lambda_p} v_p^{(r)} = \vec{b} \cdot \vec{t}_r, \quad r = 1, \ldots, m
\]

where

\[
\vec{t}_r = \frac{v_1^{(r)}}{\lambda_1} \vec{u}_1 + \cdots + \frac{v_p^{(r)}}{\lambda_p} \vec{u}_p \in \text{Col} \ U,
\]

thus the value of each unknown is the projection of \( \vec{b} \) onto each vector (46). In consequence, \( \vec{b} \in \text{Col} \ U \) guarantees the solution of (37), and it is unique only if \( p = m \).

Besides, from (42) we see that the solution \( \vec{x} = \vec{Q} \) implies that \( \vec{x} \in \text{Col} \ V \), then we have the system \( A \vec{x} = \vec{b} \) where \( \vec{x} \) and \( \vec{b} \) are totally embedded into \( \text{Col} \ V \) and \( \text{Col} \ U \), respectively, that is, \( \vec{x} \) and \( \vec{b} \) are into the subspaces of activation of \( A \), thus from (32) and (33) there is the natural inverse \( A_N^{-1} \) such that:

\[
x = A_N^{-1} \vec{b} = V_{\text{max}} \Lambda_{\text{max}}^{-1} U_{\text{max}} \vec{b} = V \Lambda^{-1} \left( \frac{b^{(1)}}{\lambda_1} \vec{u}_1 \cdots \frac{b^{(p)}}{\lambda_p} \vec{u}_p \right), \quad p = m,
\]

in accordance with (45). The vectors (46) are important because their inner products with \( \vec{b} \) give the solution of (37) via (45), and they also are remarkable because permit to construct the natural inverse:

\[
A_N^{-1}_{\text{max}} = (\vec{t}_1 \vec{t}_2 \ldots \vec{t}_m)^T, \quad p = m.
\]

Lanczos [6] considers three situations:

i). \( n < m \): The linear system is under-determined because it has more unknowns than equations, and from \( 1 \leq p \leq \min(n, m) \) is impossible the case \( p = m \), therefore, if (37) is compatible then its solution cannot be unique.

ii). \( n = m \): The system is even-determined with unique solution when \( p = m \), that is, if \( \det A \neq 0 \). In this case also \( p = n \), we have not vectors \( \vec{u}_j^{(0)} \neq \vec{0} \), thus \( \vec{b} \in \text{Col} \ U \) and automatically the system is compatible.

iii). \( n > m \): The linear system is over-determined, and by \( 1 \leq p \leq \min(n, m) \) can occur the case \( p = m \) for unique solution if the system is compatible.

Thus it is immediate the classification of linear systems introduced by Lanczos [6]:

Free and complete: \( p = n = m \), unique solution,

Restricted and complete: \( p = m < n \), over-determined, unique solution,
Free and incomplete: $p = n < m$, under-determined, non-unique solution,
Restricted and incomplete: $p < n$ and $p < m$, solution without uniqueness,
with the meanings:
Free: The conditions (30) are satisfied trivially.
Restricted: It is necessary to verify that $\vec{b} \in \text{Col} \ U$.
When $p \neq m$, the homogeneous system $A\vec{x} = \vec{0}$ has the non-trivial solutions $\vec{x}_k^{(0)}$, then from (27) we conclude that the general solution of (37) is:
$$\vec{x} = \vec{Q} + c_1 \vec{x}_1^{(0)} + \cdots + c_m-p \vec{x}_{m-p}^{(0)}$$
where the $c_k$ are arbitrary constants.

4. Conclusions

With the SVD we can find the subspaces of activation of $A$, and it leads to natural inverse [6, 26-28] of any matrix, known it in the literature as the Moore-Penrose pseudoinverse. Besides, the SVD gives a better understanding of the compatibility of linear systems. On the other hand, Lanczos [6] showed that the Singular Value Decomposition provides a universal platform to study linear differential and integral operators for arbitrary boundary conditions.

References:

[7]. F. Smithies, Linear differential operators, The Mathematical Gazette 47 (1963) 265-266
[8]. E. Schmidt, Zur theorie der linearen und nichtlinearen integralgleichungen, Teil 1, Mathematische Annalen Bd. 63 (1907) 433-476 and 64 (1907) 161-174
[9]. C. Jordan, Mémoire sur les forms bilinéaires, J. de Mathématiques Pures et Appliquées, Deuxième Série 19 (1874) 35-54
[13]. J. J. Sylvester, On the reduction of a bilinear quantic of the nth order to the form of a sum of n products, Messenger of Mathematics 19 (1889) 42-46
[16]. C. Eckart, G. Young, A principal axis transformation for non-hermitian matrices, Bull. Amer. Math. Soc. 45, No. 2 (1939) 118-121
[18]. I. J. Good, Some applications of the singular decomposition of a matrix, Technometrics 11, No. 4 (1969) 823-831
[22]. G. Golub, Aspects of scientific computing, Johann Bernoulli Lecture, Univ. of Groningen, 8th April 1996