The Singular Value Decomposition

Jody S. Hourigan and Lynn V. McIndoo
Linear Algebra- Math 45

Abstract. We explore the derivation of the SVD and its role in digital image processing. By using MATLAB, we will demonstrate how the SVD is used to minimize the size needed to store an image.

Introduction

The singular value decomposition is a highlight of linear algebra. It plays an interesting, fundamental role in many different applications, namely in digital image processing. The beauty of the SVD within its digital applications is that it provides a robust method of storing large images as smaller, more manageable square ones. This is accomplished by reproducing the original image with each succeeding nonzero singular value. Furthermore, to reduce storage size even further, one may approximate a "good enough" image with using even fewer singular values.

What is the Singular Value Decomposition?

The singular value decomposition of a matrix factors an *m x n* matrix *A* into the form

$$A = U\Sigma V^T \qquad (1)$$

where *U* is an *m x m* orthogonal matrix; *V* an *n x n* orthogonal matrix, and Σ an *m x n* matrix containing the singular values of *A*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

along its main diagonal.

A similar technique, known as the eigenvalue decomposition, diagonalizes matrix *A*, but with this case, *A* must be a square matrix. The EVD diagonalizes *A* as

$$A = VDV^{-1} \qquad (2)$$

where *D* is a diagonal matrix comprised of the eigenvalues, and *V* is a matrix whose columns contain the corresponding eigenvectors.

Unfortunately, the EVD is not always possible. What do we do then? The answer is simple: apply the SVD.

How do we know the SVD works?

Let us begin by showing that the SVD of *A* is always possible, unlike that of the EVD. Let *A* be an *m x n* matrix. The matrix A^TA is symmetric and can be diagonalized. Working with the symmetric matrix A^TA, we know two things to be true:

1. The eigenvalues of A^TA will be real and nonnegative.
2. The eigenvectors will be orthogonal.

How do we find two orthogonal matrices *U* and *V* that diagonalize a *m x n* matrix *A*? First, if the intent is to factor *A* as

$$A = U\Sigma V^T$$

then the following must be true.

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T)$$
$$A^TA = V\Sigma^T U^T U\Sigma V^T$$
$$A^TA = V\Sigma^T\Sigma V^T$$
$$A^TA = V\Sigma^2 V^T$$

Note that this implies that Σ^2 contains the eigenvalues of A^TA and V contains the corresponding eigenvectors.We now know how to find the V of the singular value decomposition $A = U\Sigma V^T$.

Next, rearrange the eigenvalues of A^TA in order of decreasing magnitude.

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = ... = \lambda_n = 0$$

Note that some eigenvalues are equal to zero.

Define the singular values of A as the square root of the corresponding eigenvalues of the matrix A^TA; that is,

$$\sigma_j = \sqrt{\lambda_j}\,,\ \text{where } j = 1, 2, ...n \tag{3}$$

Rearrange the eigenvectors of A^TA in the same order as their respective eigenvalues to produce the matrix

$$V = [v_1, v_2, ..., v_r, v_{r+1}, v_{r+2}, ..., v_n]$$

Theorem (1) *Let the rank of A be equal to r. Then r is also the rank of A^TA which is also equal to the number of nonzero eigenvalues. (Proof lies in the appendix.)*

Let $\sigma_j = \sqrt{\lambda_j}$

$V = [v_{1,...}v_{r]}$ be the set of eigenvectors associated with the non-zero eigenvalues and $V_2 = [v_{r+1}, ..., v_n]$ be the set of eigenvectors associated with zero eigenvalues.

It follows that:

$$AV_2 = (Av_{r+1}, Av_{r+2}, ..., Av_n)$$
$$AV_2 = (0, 0, ..., 0)$$
$$AV_2 = 0 \text{ where this zero is the zero matrix} \tag{4}$$

How do we find the matrix Σ? Earlier we defined the matrix Σ to be the diagonal matrix with the singular values of A along its main diagonal. From equation (3), each zero eigenvalue will result in a singular value equal to zero. Let Σ_1 be a square $r\ x\ r$ matrix containing the nonzero singular values $\{\sigma_1, \sigma_2, ...\sigma_r\}$ of A along its main diagonal. Therefore matrix Σ may be represented by:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

where the singular values along the diagonal are arranged in decreasing magnitude, and the zero singular values are placed at the end of the diagonal. This new matrix Σ, with the correct dimension $m\ x\ n$, is padded with $m - r$ rows of zeros and $n - r$ columns of zeros.

But how do we find the orthogonal matrix U? Looking at the equation

$$AV = U\Sigma$$

it follows that:

$$A[v_1, ..., v_r, v_{r+1}, ...v_n] = [u_1, ...u_r, u_{r+1}, ...u_n] \begin{bmatrix} \sigma_1 & & & & & \\ & ... & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & ... & \\ & & & & & 0 \end{bmatrix}$$

$$[Av_1, ...Av_r, Av_{r+1}, ..., Av_n] = [\sigma_1 u_1, ...\sigma_r u_r, 0...0]$$

Therefore,

$$\text{for } j = 1, ..., r$$

$$Av_j = \sigma_j u_j \tag{5}$$

Examining equation (5), let us remember that σ_j is a scalar and that v_j, and u_j are column vectors and that matrix vector multiplication results in another vector. Therefore, the vector resulting from the multiplication of Av_j is equal to the vector u_j multiplied by the scalar σ_j. We can look at the vector Av_j as lying in the direction of the unit vector u_j with absolute length σ_j. Av_j can be calculated because we previously found matrix *V*. Therefore, the unit vector u_j, is a result of dividing the vector Av_j by its magnitude, σ_j.

$$u_j = \frac{Av_j}{\sigma_j} \tag{6}$$

Equation (6) is restricted to the first *r* nonzero singular values. (We will explain the case for singular values equal to zero later.) This allows us to find the column vectors of *U* so long as there is no division by zero. Therefore we see that this method allows us to find only part of the matrix *U*. How do we find the other part where the singular values of *A* are equal to zero? As we saw before in the matrix *V*, the matrix *U* may be defined as:

$$U = [U_1, U_2]$$

Letting

$$U_1 = [u_1, ..., u_r]$$

$$U_2 = [u_{r+1}, ..., u_m]$$

Then

$$
\begin{aligned}
AV_1 &= A[v_1, ..., v_2] \\
&= [Av_1, ..., Av_r] \\
&= [\sigma_1 u_1, ..., \sigma_r u_r] && \text{From equation (5)} \\
&= [u_1, ..., u_r] \begin{bmatrix} \sigma_1 & & \\ & ... & \\ & & \sigma_r \end{bmatrix} \\
AV_1 &= U_1 \Sigma_1 && (7)
\end{aligned}
$$

Before we proceed to find the matrix U_2, let us pause for a minute. Referring to the illustration of the four fundamental subspaces (see appendix) we remember that the nullspace $N(A)$, of a matrix A, denotes the set of all nontrivial (non-zero) solutions to equation $Ax = 0$. Using equation (4)

$$AV_2 = 0 \text{ where zero represents the zero matrix}$$

it follows that V_2 forms a basis for the $N(A)$. Also because

$$
\begin{aligned}
Av_j &= 0u_j \quad \text{where } j = r+1, r+2, ..., n \\
Av_j &= 0 \\
v_j &\in N(A) && (7)
\end{aligned}
$$

We know that the orthogonal complement to the $N(A)$ is the $R(A^T)$. Since the columns in the matrix V are orthogonal, the remaining vectors $v_1 ... v_r$ must lie in the subspace corresponding to the $R(A^T)$. Let us now return to finding U. From equation 6, we see that $u_n = \frac{1}{\sigma_n} Av_n$. This equation holds the valuable information that the column vectors of U, $[u_1, ..., u_r]$ are in the columnspace of A. This is because the column vectors of U are linear combinations of the columns of A. Or, in matrix notation

$$u_j \in R(A) \quad \text{where} \quad j = r+1, r+2, ..., n$$

It now follows that $R(A)$ and $N(A^T)$ are orthogonal complements. Since the matrix U is an orthogonal matrix and the first r column vectors of U have been assigned to lie in the $R(A)$, then $[u_{r+1}, ... u_m]$ must lie in the $N(A^T)$. This is a very exciting discovery! The vectors that lie in the $N(A^T)$ are the vectors $[u_{r+1}, ..., u_m]$ which form the matrix U_2. Now that we have the matrix V, the matrix Σ, and the matrix U, the singular value decomposition has been found for any matrix A. Lets see if $U \Sigma V^T$ actually does diagonalize and equal the matrix A.

$$
\begin{aligned}
U\Sigma V^T &= [U_1, U_2]\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}[V_1 V_2]^T \\
&= [U_1, U_2]\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \\
&= [U_1, U_2]\begin{bmatrix} \Sigma_1 V_1^T \\ 0 \end{bmatrix} \\
&= U_1 \Sigma_1 V_1^T + 0 \\
&= U_1 \Sigma_1 V_1^T && \text{From equation (7)} \\
&= A V_1 V_1^T \\
&= AI \\
&= A \\
U\Sigma V^T &= A
\end{aligned}
$$

The singular value decomposition really does diagonalize *A*!

MATLAB'S Contribution to the Singular Value Decomposition

MATLAB plays an important role in allowing students the opportunity to better understand and visualize how the SVD works. The SVD by hand would be awkward and tedious for anything other than very small matrices. But with MATLAB, students are able to deal with any size image and may even write their own programs to see the SVD in action. For example, here is an m-file which applies the SVD to an image with dimension 500 x 337.

```
close all
I=imread('sanfran.tif');
imshow(I);
J=im2double(I);
figure
imshow(J)
[U,S,V]=svd(J);
for k = 1:10:100
 K=U(:,1:k)*S(1:k,1:k)*V(:,1:k)';
 imshow(K)

 pause
 end
```

To understand how the SVD is used to approximate an image or obtain an exact representation, let us first take another look at how the SVD breaks down the matrix *A* into three separate matrices.

$$\begin{aligned} A &= U\Sigma V^T \\ &= [u_1, \ldots u_n] \begin{bmatrix} \sigma_1 & & \\ & \ldots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ : \\ v_n^T \end{bmatrix} \\ &= [u_1, \ldots u_n] \begin{bmatrix} \sigma_1 v_1^T \\ : \\ \sigma_n v_n^T \end{bmatrix} \\ &= \sigma_1 u_1 v_1^T + \ldots + \sigma_n u_n v_n^T \\ &= \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T \text{ because } \sigma_{r+1} \ldots \sigma_n = \text{are all equal to zeros} \end{aligned}$$

This m-file was created to approximate the 500 x 337 image by using the first 100 singular values in steps of ten. For example, the first singular value provides a rough approximation by

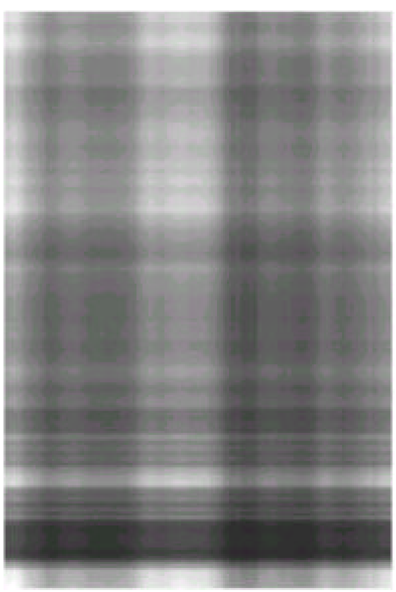
$$A = \sigma_1 u_1 v_1^T$$

The next rough, but much better approximation, (keeping in mind the m-file is incremented in steps of ten) is given by

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_{10} u_{10} v_{10}^T$$

And on and on, until finally we see that we don't need even need to go up to the 100th singular value because a "good enough" approximation to the original image may be given by the 70th singular value such that:

$$A = \sigma_1 u_1 v_1^T + \ldots + \sigma_{70} u_{70} v_{70}^T$$



$A = \sigma_1 u_1 v_1^T$

$A = \sigma_1 u_1 v_1^T + \ldots + \sigma_{10} u_{10} v_{10}^T$

$A = \sigma_1 u_1 v_1^T + \ldots + \sigma_{30} u_{30} v_{30}^T$

$A = \sigma_1 u_1 v_1^T + ... + \sigma_{50} u_{50} v_{50}^T$

$A = \sigma_1 u_1 v_1^T + ... + \sigma_{70} u_{70} v_{70}^T$

The original image

The original image has a full rank of 337. Therefore, an exact representation may be given by

$$A = \sigma_1 u_1 v_1^T + ... + \sigma_r u_r v_r^T \quad \text{where } r = 337$$

However, by comparing the last two approximations, it is clear that a “good enough” representation may be found with far fewer singular values. This substantially reduces the amount of information necessary to store the exact image. This is a real life example of the power and beauty of the singular value decomposition. Let us look at the numbers involved. Because the original image before the SVD was 500 x 337, it required 168,500 entries for storage. After the SVD was applied, the original image needed 113,569 entries. This image requires 67.4 % of the original for an exact representation. However, as we saw above, a “good enough” image may be produced using far fewer entries. The image produced by the first 50 singular values requires only 2,500 entries; the image with the first 70 requires 4,900. This drastically reduces the information necessary–to 1.5% and 2.9% of the original image, respectively. Pretty fantastic results!!!

Conclusion

As we have seen, the singular value decomposition is pretty special in that it can work its magic on virtually any matrix; whether it is square, rectangular, invertible, or not. Further, after the SVD has been applied, it allows for a substantial reduction in the amount of storage required, making large images more manageable and easier to work with.

Appendix

Proof of theorem 1. If

$$x \in N(A)$$

then we know the following to be true:

$$Ax = 0$$

Multiplying by A^T

$$
\begin{aligned}
A^T A x &= A^T 0 \\
A^T A x &= 0 \\
x &\in N(A^T A)
\end{aligned}
$$

Therefore

$$N(A) \subset N(A^T A)$$

If

$$x \in N(A^T)$$

Then we know the following to be true:

$$A^T A x = 0$$

Multiplying by x^T

$$
\begin{aligned}
x^T A^T A x &= x^T \\
(Ax)^T A x &= 0 \\
\|Ax\|^2 &= 0 \\
\|Ax\| &= 0 \\
Ax &= 0 \\
x &\in N(A)
\end{aligned}
$$

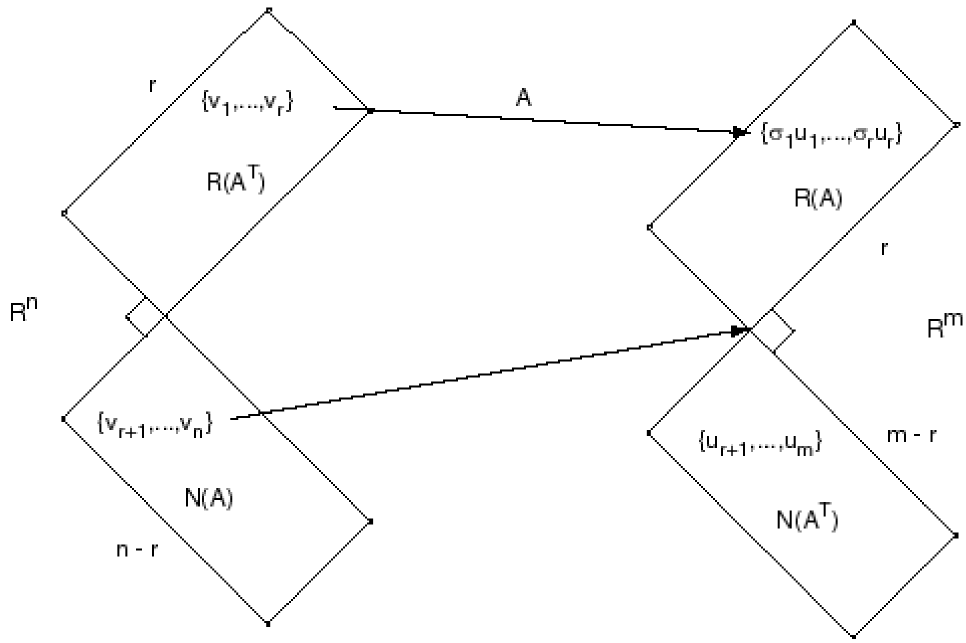
Therefore

$$
\begin{aligned}
N(A^T A) &\subset N(A) \\
N(A^T A) &= N(A) \qquad (8)
\end{aligned}
$$

Let A have dimension *mxn*. $A^T A$ has dimension *nxn*. It follows from (8) that the nullity of $(A^T A)$ = nullity of (A) and is equal to k, and

$$
\begin{aligned}
N(A^T A) &= N(A) = k \\
rank(A^T A) &= n - k \\
rank(A) &= n - k \\
rank\ of\ A &= rank\ of\ A^T A
\end{aligned}
$$

To explain why the rank r is equal to the number of nonzero eigenvalues first look at the illustration of the four fundamental subspaces. Remember that $[v_1, \ldots, v_r]$ are the eigenvectors corresponding to the nonzero eigenvalues of A, and that the remaining column vectors of V correlate with the eigenvalues equal to zero. So there exists r nonzero singular values. Thus r is equal to the number of nonzero eigenvalues.



The Four Fundamental Subspaces

Bibliography

Lay, David C. **Linear Algebra and Its Applications**. Addison Wesley Longman, Inc., 1997

Leon, Steven. **Atlast**:**Computer Exercises for Linear Algebra**. Prentice-Hall, Inc., 1996

Leon, Steven. **Linear Algebra with Applications**. Prentice Hall, Inc., 1998

Mulcahy, Colm & Rossi, John. "Atlast: A Fresh Approach to Singular Value Decomposition"

Strang, Gilbert. **Introduction to Linear Algebra** Wellesley-Canbridge Press, 1998