8.3 Singular Values

Example 1 Show that if $L(\vec{x}) = A\vec{x}$ is a linear transformation from R^2 to R^2, then there are two orghogonal unit vectors $\vec{v}_1$ and $\vec{v}_2$ in R^2 such that $L(\vec{v}_1)$ and $L(\vec{v}_2)$ are orthogonal as well.

Solution This statement is clear for some classes of transformation, for example,

1. If L is an orthogonal transformation, then any two orghogonal unit vectors $\vec{v}_1$ and $\vec{v}_2$ will do, by Fact 5.3.2.

2. If A is symmetric, then we can choose two orthogonal unit eigenvectors, by the spectral theorem.

However, for an arbitrary linear transformation L, the statement isn't that obvious.

Hint: Consider an orthonormal eigenbasis $\vec{v}_1$, $\vec{v}_2$ of the symmetric matrix A^TA, with associated eigenvalues λ_1, λ_2. $L(\vec{v}_1) = A\vec{v}_1$ and $L(\vec{v}_2) = A\vec{v}_2$ are orthogonal, as claimed:

$$(A\vec{v}_1) \cdot (A\vec{v}_2) = (A\vec{v}_1)^T A\vec{v}_2 = \vec{v}_1^T A^T A\vec{v}_2$$

$$= \vec{v}_1^T(\lambda_2\vec{v}_2) = \lambda_2(\vec{v}_1 \cdot \vec{v}_2) = 0$$

Note that $\vec{v}_1$, $\vec{v}_2$ need not be eigenvectors of matrix A.

Example 2 Consider the linear transformation $\vec{x}) = A\vec{x}$, where $A = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix}$.

1. Find an orthonormal basis $\vec{v}_1$, $\vec{v}_2$ of R^2 such that $L(\vec{v}_1)$ and $L(\vec{v}_2)$ are orthogonal.

2. Show that the image of the unit circle under transformation L is an ellipse. Find the lengths of the two semiaxes of this ellipse, in terms of the eigenvalues of matrix A^TA.

Solution

1. Using the ideas of Example 1

$$A^TA = \begin{bmatrix} 6 & -7 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix} = \begin{bmatrix} 85 & -30 \\ -30 & 40 \end{bmatrix}$$

The characteristic polynormial of A^TA is

$$\lambda^2 - 125\lambda + 2500 = (\lambda - 100)(\lambda - 25),$$

so the corresponding eigenspaces are

$$E_{100} = ker \begin{bmatrix} 15 & 30 \\ 30 & 60 \end{bmatrix} = span \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

$$E_{25} = ker \begin{bmatrix} -60 & 30 \\ 30 & -15 \end{bmatrix} = span \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

For orthonormal basis

$$\vec{v}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \vec{v}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

2. The unit circle consists of the form $\vec{x} = \cos(t)\vec{v}_1 + \sin(t)\vec{v}_2$, and the image of the unit circle consists of the form

$$L(\vec{x}) = \cos(t)L(\vec{v}_1) + \sin(t)L(\vec{v}_2)$$

The image is the ellipse whose semimajor and seminor axes are $||L(\vec{v}_1)||$ and $||L(\vec{v}_2)||$:

$$||L(\vec{v}_1)||^2 = (A\vec{v}_1)(A\vec{v}_1) = \vec{v}_1^T A^T A\vec{v}_1 = \vec{v}_1^T(\lambda_1\vec{v}_1) = \lambda_1$$

Likewise,

$$||L(\vec{v}_2)||^2 = \lambda_2.$$

Thus

$$||L(\vec{v}_1)|| = \sqrt{\lambda_1} = \sqrt{100} = 10$$

$$||L(\vec{v}_2)|| = \sqrt{\lambda_2} = \sqrt{25} = 5$$

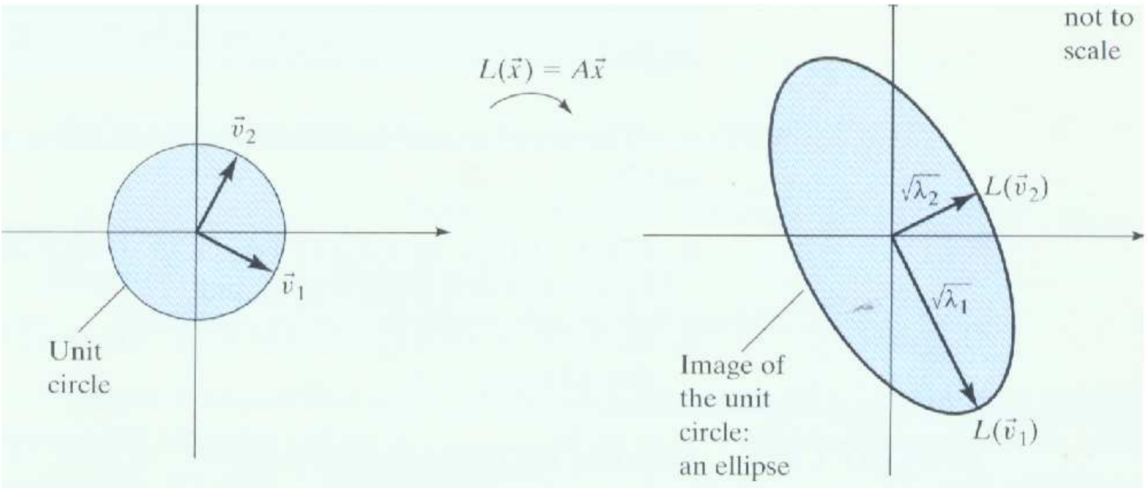
We can also compute $L(\vec{v}_1)$ and $L(\vec{v}_2)$ directly:

$$L(\vec{v}_1) = A\vec{v}_1 = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 10 \\ -20 \end{bmatrix}$$

$$L(\vec{v}_2) = A\vec{v}_2 = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 10 \\ 5 \end{bmatrix}$$

So that

$$||L(\vec{v}_1)|| = 10, ||L(\vec{v}_2)|| = 5$$



Definition 8.3.1 Singular values
The `singular values` of an $m \times n$ matrix A are the *square roots* of the eigenvalues of the symmetric $n \times n$ matrix $A^T A$, listed with their algebraic multiplicities. It is customary to denote the singular values by $\sigma_1, \sigma_2, \ldots, \sigma_n$, and to list them in decreasing order:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$$

Fact 8.3.2 The image of the unit circle
Let $L(\vec{x}) = A\vec{x}$ be an invertible linear transformation from R^2 to R^2. The image of the unit circle under L is an ellipse E. The lengths of the semimajor and the seminor axes of E are the singular values σ_1, and σ_2 of A, respectively.

Fact 8.3.3
Let $L(\vec{x}) = A\vec{x}$ be a linear transformation from R^n to R^m. Then there is an orghonormal basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ of R^n such that

1. vectors $L(\vec{v}_1), L(\vec{v}_2), \ldots, L(\vec{v}_n)$ are orthogonal, and

2. the lengths of these vectors are the singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$ of matrix A.

To construct $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, find an orthonormal eigenbasis for matrix A^TA. Make sure that the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ appear in descending order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

Proof

1. $L(\vec{v}_i) \cdot L(\vec{v}_j) = (A\vec{v}_i) \cdot (A\vec{v}_j) = (A\vec{v}_i)^T A\vec{v}_j$ $= \vec{v}_i^T A^T A\vec{v}_j = \vec{v}_i^T(\lambda_j \vec{v}_j) = \lambda_j(\vec{v}_i \cdot \vec{v}_j) = 0$ when $i \neq j$, and

2. $||L(\vec{v}_i)||^2 = (A\vec{v}_i) \cdot (A\vec{v}_i) = \vec{v}_i^T A^T A\vec{v}_i$ $= \vec{v}_i^T(\lambda_i \vec{v}_i) = \lambda_i(\vec{v}_i \cdot \vec{v}_i) = \lambda_i = \sigma_i^2 \geq 0$, so that $||L(\vec{v}_i)|| = \sigma_i$.

Example 3 Consider the linear transformation

$$L(\vec{x}) = A\vec{x}, A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

a. Find the singular values of A.
b. Find orthonormal vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, in R^3 such that $L(\vec{v}_1), L(\vec{v}_2), L(\vec{v}_3)$ are orthogonal.
c. Sketch and describe the image of the unit sphere under the transformation L.

Solution

a.

$$A^TA = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

The eigenvalues are $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$. The singular values of A are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}, \sigma_2 = \sqrt{\lambda_2} = 1, \sigma_3 = \sqrt{\lambda_3} = 0$$

b. Find an orthonormal eigenbasis $\vec{v}_1, \vec{v}_2, \vec{v}_3$, for $A^T A$:

$$E_3 = span \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, E_1 = span \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, E_0 = span \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \vec{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Compute $L(\vec{v}_1), L(\vec{v}_2), L(\vec{v}_3)$ and check orthogonality:

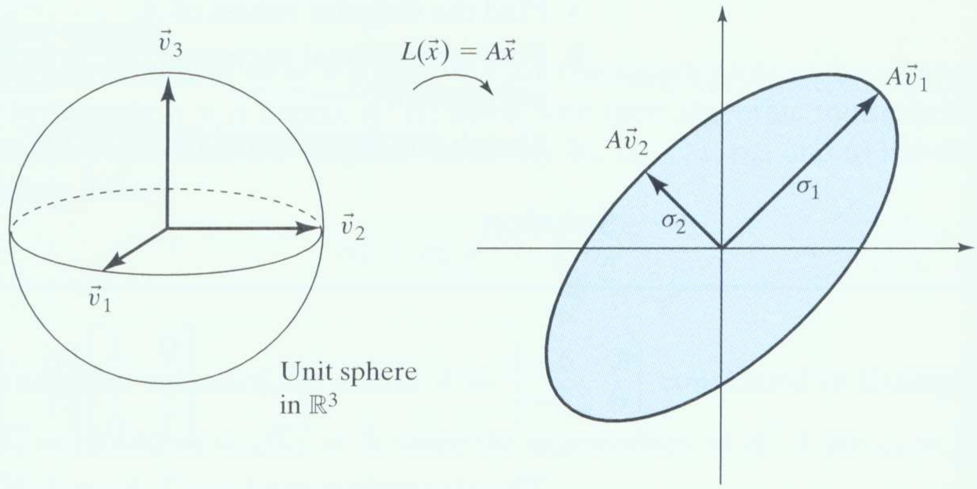
$$A\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, A\vec{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}, A\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

c. The unit sphere in R^3 consists of all vectors of the form $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3$, where $c_1^2 + c_2^2 + c_3^2 = 1$.

The image of the unit sphere consists of the vectors

$$L(\vec{x}) = c_1 L(\vec{v}_1) + c_2 L(\vec{v}_2)$$

where $c_1^2 + c_2^2 \leq 1$. The image is the full ellipse shaded in Figure 3.



$\vec{v}_3$
$L(\vec{x}) = A\vec{x}$
$A\vec{v}_1$
$A\vec{v}_2$
σ_1
σ_2
$\vec{v}_2$
$\vec{v}_1$
Unit sphere
in $\mathbb{R}^3$

Figure 3

Example 3 shows that some of the singular values of a matrix may be zero. Suppose the singular values $\sigma_1, \sigma_2, \ldots, \sigma_s$ are nonzero, while $\sigma_{s+1}, \sigma_{s+2}, \ldots, \sigma_n$ are zero. Choose eigenbasis $\vec{v}_1, \ldots, \vec{v}_s, \vec{v}_{s+1}, \ldots, \vec{v}_n$ of $A^T A$ for R^n. Note that $||A\vec{v}_i|| = \sigma_i = 0$ and therefore $A\vec{v}_i = \vec{0}$ for $i = s+1, \ldots, n$.

We claim that the vectors $A\vec{v}_1, \ldots, A\vec{v}_s$ form a basis of the image of A, since any vector in the image of A can be written as

$$\begin{aligned} A\vec{x} &= A(c_1\vec{v}_1 + \ldots + c_s\vec{v}_s + \ldots + c_n\vec{v}_n) \\ &= c_1 A\vec{v}_1 + \ldots + c_s A\vec{v}_s \end{aligned}$$

This shows that $s = dim(imA) = rank(A)$.

Fact 8.3.4

If A is an $m \times n$ matrix of rank r, then the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ are nonzero, while $\sigma_{r+1}, \ldots, \sigma_n$ are zero.

Singular Value Decomposition

Fact 8.3.3 can be expressed in terms of a matrix decomposition.

Consider a linear transformation $L(\vec{x}) = A\vec{x}$ from R^n to R^m, and choose an orthonormal basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ as in Fact 8.3.3. Let $r = rank(A)$. We know that the vectors
$A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r$ are orthogonal and nonzero, with $||A\vec{v}|| = \sigma_i$. We introduce the unit vectors

$$\vec{u}_1 = \frac{1}{\sigma_1} A\vec{v}_1, \ldots, \vec{u}_r = \frac{1}{\sigma_r} A\vec{v}_r$$

We can write

$$A\vec{v}_i = \sigma_i \vec{u}_i \text{ for } i = 1, 2, \ldots, r$$

and

$$A\vec{v}_i = \vec{0} \text{ for } i = r+1, r+2, \ldots, n$$

We can express these equations in matrix form as follows:

$$A \underbrace{\begin{bmatrix} | & & | & | & & | \\ \vec{v}_1 & \dots & \vec{v}_r & \vec{v}_{r+1} & \dots & \vec{v}_n \\ | & & | & | & & | \end{bmatrix}}_{V}$$

$$= \begin{bmatrix} | & & | & | & & | \\ \sigma_1\vec{u}_1 & \dots & \sigma_r\vec{u}_r & \vec{0} & \dots & \vec{0} \\ | & & | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & & | & | & & | \\ \vec{u}_1 & \dots & \vec{u}_r & \vec{0} & \dots & \vec{0} \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ & 0 & & 0 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} | & & | & | & & | \\ \vec{u}_1 & \dots & \vec{u}_r & \vec{u}_{r+1} & \dots & \vec{u}_m \\ | & & | & | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ & 0 & & 0 \end{bmatrix}}_{\Sigma}$$

The vector space $ker(A^T)$ has dimesion $m - r$. Let $\{\vec{u}_{r+1}, \vec{u}_{r+2}, \dots, \vec{u}_m\}$ be an orthonormal basis for $ker(A^T)$. Then $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m$ form an orthonormal basis for R^m.

Note that V is an orthogonal $n \times n$ matrix, U is an orthogonal $m \times m$ matrix, and Σ is an $m \times n$ matrix whose first r diagonal entries are $\sigma_1, \sigma_2, \ldots, \sigma_r$, and all other entries are zero.

Fact 8.3.5 Singular-value decomposition

Any $m \times n$ matrix A can be written as

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

where U is an orthogonal $m \times m$ matrix; V is an orthogonal $n \times n$ matrix; and Σ is an $m \times n$ matrix whose first r diagonal entries are the nonzero sigular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of A, and all other entries are zero (where $r = rank(A)$).

Alternatively, this singular value decomposition can be written as

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_r \vec{u}_r \vec{v}_r^T,$$

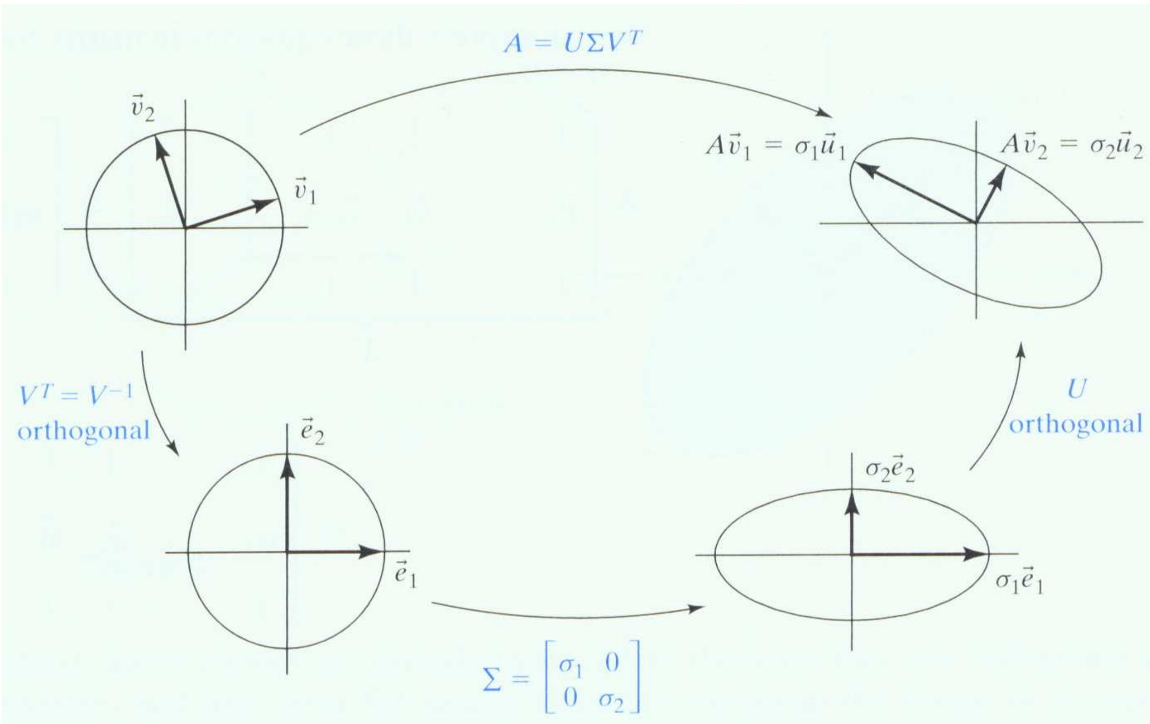
where $\vec{u}_i$ and $\vec{v}_i$ are the columns of U and V, respectively.

Proof

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

$$= \begin{bmatrix} \vec{u}_1 & \ldots \vec{u}_r & \ldots \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \ddots & \\ 0 & & & & 0 \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_r^T \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vec{u}_1 & \ldots \vec{u}_r & \ldots \end{bmatrix} \begin{bmatrix} \sigma_1 \vec{v}_1^T \\ \vdots \\ \sigma_r \vec{v}_r^T \\ \vdots \end{bmatrix} = \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_r \vec{u}_r \vec{v}_r^T$$



Consider a singular value decomposition $A = U\Sigma V^T$, where

$$V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_n \\ | & & | \end{bmatrix} \text{ and } U = \begin{bmatrix} | & & | \\ \vec{u}_1 & \dots & \vec{u}_m \\ | & & | \end{bmatrix}$$

We know that

$$A\vec{v}_i = \sigma_i \vec{u}_i \quad for \quad i = 1, 2, \dots, r$$

and

$$A\vec{v}_i = \vec{0} \quad for \quad i = r + 1, \dots, n$$

These equations tell us that

$$im(A) = span(\vec{u}_1, \dots, \vec{u}_r)$$

and

$$ker(A) = span(\vec{v}_{r+1}, \dots, \vec{v}_n)$$

That is, SVD provides us with orthonormal bases for the kernel and image of A.

Likewise, we have $A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$ or $A^T U = V\Sigma^T$.
Reading the last equation column by column, we find that

$$A^T\vec{u}_i = \sigma_i\vec{v}_i \quad for \quad i = 1, 2, \ldots, r$$

and

$$A^T\vec{u}_i = \vec{0} \quad for \quad i = r+1, \ldots, m$$

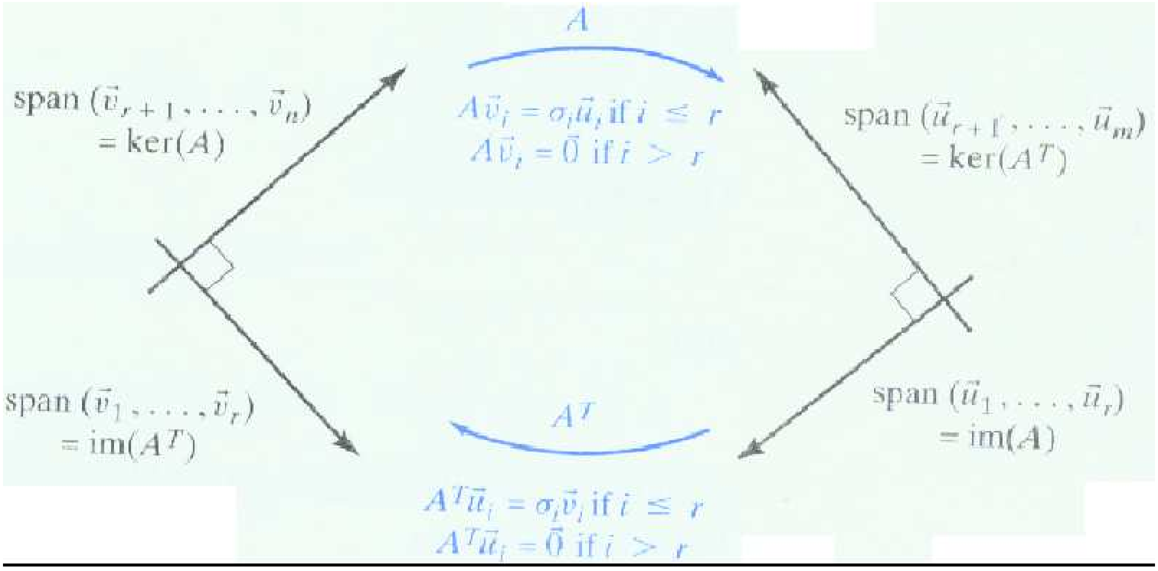
As before

$$im(A^T) = span(\vec{v}_1, \ldots, \vec{v}_r)$$

and

$$ker(A^T) = span(\vec{u}_{r+1}, \ldots, \vec{u}_m)$$

See Figure 5



$$
\begin{array}{ccccc}
 & R^n & \xrightarrow{\underline{A : m \times n}} & R^m & \\
 & \vec{v}_1 & & \vec{u}_1 & \\
im(A^T) & \vdots & & \vdots & im(A) \\
= Row(A) & \vec{v}_r & & \vec{u}_r & = Col(A) \\
---- & --- & ------ & --- & ---- \\
 & \vec{v}_{r+1} & & \vec{u}_{r+1} & \\
ker(A) & \vdots & & \vdots & ker(A^T) \\
 & \vec{v}_n & & \vec{u}_m &
\end{array}
$$

Example 5 Find an SVD for $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$

Solution

$$V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix},$$

$$U = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ -2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

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

Compare with Example 3 where $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

Example 1 Consider an $m \times n$ matrix A of rank r, and a singular value decomposition $A = U\Sigma V^T$. Explain how you can express the least-squares solutions of a system $A\vec{x} = \vec{b}$ as a linear combinations of the columns $\vec{v}_1, \ldots, \vec{v}_n$ of V.

Solution Let $\vec{x} = c_1\vec{v}_1 + \ldots + c_n\vec{v}_n$ is a least squares solution if $A\vec{x} = \sum_{i=1}^n c_i A\vec{v}_i = \sum_{i=1}^r c_i\sigma_i\vec{u}_i = proj_{imA}\vec{b}$.

We know that $proj_{imA}\vec{b} = \sum_{i=1}^r (\vec{b} \cdot \vec{u}_i)\vec{u}_i$ since $\vec{u}_1, \ldots, \vec{u}_r$ is an orthonormal basis of im(A). Comparing the coefficient of $\vec{u}_i$, we find that $c_i\sigma_i = \vec{b} \cdot \vec{u}_i$ or $c_i = \frac{\vec{b}\cdot\vec{u}_i}{\sigma_i}$, for $i = 1, \ldots, r$, while no condition is imposed on $c_{r+1}, \ldots, c_n$. Therefore, the least squares solutions are of the form

$$\vec{x}^* = \sum_{i=1}^{r} \frac{\vec{b} \cdot \vec{u}_i}{\sigma_i}\vec{v}_i + \sum_{i=r+1}^{n} c_i\vec{v}_i$$

where $c_{r+1}, \ldots, c_n$ are arbitrary.

Example 2 Consider an SVD $A = U\Sigma V^T$ of an $m \times n$ matrix A. Show that the columns of U form an orthonormal eigenbasis for AA^T. What are the associated eigenvalues? What does your answer tell you about the relationship between the eigenvalues of A^TA and AA^T.

Solution

$$AA^TU = (U\Sigma V^T)(U\Sigma V^T)^TU = U\Sigma V^TV\Sigma^TU^TU$$

$$= U\Sigma\Sigma^T$$

$$AA^T\vec{u}_i = \begin{cases} \sigma_i^2\vec{u}_i & for \quad i = 1, \ldots, r \\ \vec{0} & for \quad i = r+1, \ldots, n \end{cases}$$

The columns of U form an orthonormal eigenbasis for AA^T. The associated eigenvalues are the squares of the singular values.