Unit 4: Singular Value Decomposition

Prof.Dr.P.M.Bajracharya

School of Mathematical Sciences
T.U., Kirtipur

May 7, 2024

Summary

Eigenspaces

The set of all eigenvalues of a matrix has a special name.

Spectrum

The set of all eigenvalues of a matrix is called the **spectrum** of a matrix.

The set of all eigenvalues of a matrix has a special name.

Spectrum

The set of all eigenvalues of a matrix is called the **spectrum** of a matrix.

Eigenspace

The set of all eigenvectors corresponding to an eigenvalue of a matrix together with the zero vector is called the **Eigenspace** associated with the eigenvalue.

Theorem

The **Eigenspace** associated with an eigenvalue of a matrix is a vector space.

Theorem

The **Eigenspace** associated with an eigenvalue of a matrix is a vector space.

Proof. If v and w are eigenvectors associated with the same eigenvalue λ, then

$$A(v+w) = Av + Aw = \lambda v + \lambda w = \lambda(v+w)$$

Theorem

The **Eigenspace** associated with an eigenvalue of a matrix is a vector space.

Proof. If v and w are eigenvectors associated with the same eigenvalue λ, then

$$A(v+w) = Av + Aw = \lambda v + \lambda w = \lambda(v+w)$$

and for any realnumber c,

$$A(cv) = c(Av) = c(\lambda v) = \lambda(cv).$$

$\square$

Let A be a symmetric $n \times n$ matrix. By spectral decomposition theorem,

$$A = V \Lambda V^T, \tag{1}$$

where Λ is the diagonal matrix of eigenvalues λ_i's arranged in decreasing order and V is the $n \times n$ orthogonal matrix of the eigenvectors associated with the eigenvalues λ_i's.

From Equation 1, we get

$$A = (u_1 \ \dots \ u_n) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix}$$

From Equation 1, we get

$$A = (u_1 \; \dots \; u_n) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix}$$

$$= (\lambda_1 u_1 \; \dots \; \lambda_n u_n) \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix}$$

From Equation 1, we get

$$A = \begin{pmatrix} u_1 & \dots & u_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 u_1 & \dots & \lambda_n u_n \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix}$$

$$= \sum_{i=1}^{n} \lambda_i u_i u_i^T .$$

Thus, we have expressed the matrix A as a sum of n terms, each of which is the outer product of an eigenvector with its transpose, and is scaled by its associated eigenvalue.

Equation (1) gives the spectral decomposition of the matrix A, also called the **Eigenvalue Decomposition** (EVD).

Equation (1) gives the spectral decomposition of the matrix A, also called the **Eigenvalue Decomposition** (EVD).

The spectral decomposition theorem of a symmetric matrix A can be restated as follows:

Theorem

If an $n \times n$ matrix A is symmetric, then A is orthogonally diagonalizable.

Its convers is also true.

Theorem

If an $n \times n$ matrix A is orthogonally diagonalizable, then A is a symmetric matrix.

Theorem

If an $n \times n$ matrix A is orthogonally diagonalizable, then A is a symmetric matrix.

Proof.

If A is orthogonally diagonalizable, then

$$A^T = (PDP^T)^T = P^{TT}D^TP^T = PDP^T = A.$$

Thus, A is symmetric. $\square$

Singular Value Decomposition

We have seen that symmetric matrices are always (orthogonally) diagonalizable. That is, for any symmetric matrix $A \in \mathbb{R}^{n\times n}$, there exist an orthogonal matrix $V = (v_1 \ \dots \ v_n)$ and a diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, both real and square, such that

$$A = V\Lambda V^T.$$

We have seen that symmetric matrices are always (orthogonally) diagonalizable. That is, for any symmetric matrix $A \in \mathbb{R}^{n\times n}$, there exist an orthogonal matrix $V = (v_1 \ \ldots \ v_n)$ and a diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, both real and square, such that

$$A = V\Lambda V^T.$$

What about general rectangular matrices?

We have seen that symmetric matrices are always (orthogonally) diagonalizable. That is, for any symmetric matrix $A \in \mathbb{R}^{n\times n}$, there exist an orthogonal matrix $V = (v_1 \ \dots \ v_n)$ and a diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, both real and square, such that

$$A = V\Lambda V^T.$$

What about general rectangular matrices?

Let's now consider the generalization of the spectral theorem to an arbitrary $m \times n$ matrix A.

The singular value decomposition is based on the following property of the ordinary diagonalization that can be imitated for rectangular matrices:

The singular value decomposition is based on the following property of the ordinary diagonalization that can be imitated for rectangular matrices: The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks certain vectors (the eigenvectors).

If $Ax = \lambda x$ and $\|x\| = 1$, then

$$\|Ax\| = \|\lambda x\| = |\lambda|\|x\| = |\lambda|.$$

If λ_1 is the largest eigenvalue, then a corresponding unit eigenvector v_1 identifies a direction in which the stretching effect of A is greatest.

If λ_1 is the largest eigenvalue, then a corresponding unit eigenvector v_1 identifies a direction in which the stretching effect of A is greatest. That is, the length of Ax is maximized when

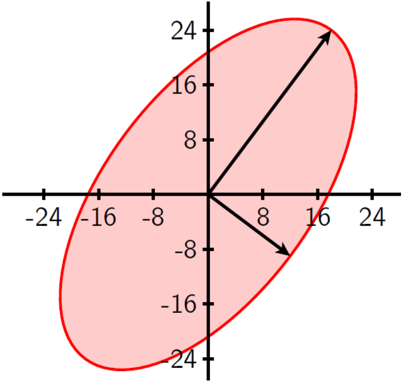
$$x = v_1, \text{ and } \|Av_1\| = |\lambda_1|.$$

If λ_1 is the largest eigenvalue, then a corresponding unit eigenvector v_1 identifies a direction in which the stretching effect of A is greatest. That is, the length of Ax is maximized when

$$x = v_1, \text{ and } \|Av_1\| = |\lambda_1|.$$

Example

If $A = \begin{pmatrix} 2 & 1 \\ 2 & 5 \end{pmatrix}$, then the linear transformation $x \mapsto Ax$ maps the unit circle $\{x : \|x\| = 1\}$ in $\mathbb{R}^2$ onto an ellipse in $\mathbb{R}^2$, shown in the figure. Find a unit vector x at which the length $\|Ax\|$ is maximized, and compute this maximum length.


24
16
8
-8
-16
-24
-24
-16
-8
8
16
24

Solution. The quantity $\|Ax\|^2$ is maximized at the same x that maximizes $\|Ax\|$, and$\|Ax\|^2$ is easier to study. Observe that

$$\|Ax\|^2 = (Ax)^T(Ax) = x^TA^TAx = x^T(A^TA)x.$$

Also, A^TA is a symmetric matrix, since

$$(A^TA)^T = A^TA^{TT} = A^TA.$$

So the problem now is

$$\begin{array}{ll} \text{to maximize:} & x^T(A^TA)x \\ \text{subject to:} & \|x\| = 1. \end{array}$$

We know that the maximum value is the greatest eigenvalue λ_1 of A^TA. Also, the maximum value is attained at a unit eigenvector of A^TA corresponding to λ_1.

For the matrix A in this example,

$$A^T A = \begin{pmatrix} 2 & 2 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 8 & 12 \\ 12 & 26 \end{pmatrix}.$$

The eigenvalues of $A^T A$ are

$$\lambda_1 = 32, \lambda_2 = 2.$$

Corresponding unit eigenvectors are, respectively,

$$v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

The maximum value of $\|Ax\|^2$ is 32, attained when x is the unit vector v_1. The vector Av_1 is a point on the ellipse in the above figure farthest from the origin, namely,

$$Av_1 = \begin{pmatrix} 2 & 1 \\ 2 & 5 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 4 \\ 12 \end{pmatrix}.$$

Therefore, for $\|x\| = 1$, the maximum value of $\|Ax\|$ is

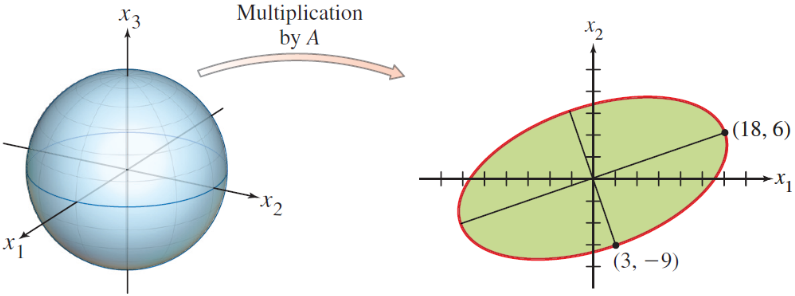
$$\|Av_1\| = \sqrt{32} = 4\sqrt{2}.$$

$\square$

This description of v_1 and $|\lambda_1|$ has an analogue for rectangular matrices that will lead to the singular value decomposition.

Example

If $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$, then the linear transformation $x \mapsto Ax$ maps the unit sphere $\{x : \|x\| = 1\}$ in $\mathbb{R}^3$ onto an ellipse in $\mathbb{R}^2$, shown in the figure. Find a unit vector x at which direction the length $\|Ax\|$ is maximized, and compute this maximum length.


x_3
x_2
x_1
Multiplication
by A
x_2
(18, 6)
x_1
(3, −9)

Solution. The quantity $\|Ax\|^2$ is maximized at the same x that maximizes $\|Ax\|$, and $\|Ax\|^2$ is easier to study.

Solution. The quantity $\|Ax\|^2$ is maximized at the same x that maximizes $\|Ax\|$, and $\|Ax\|^2$ is easier to study. Observe that

$$\|Ax\|^2 = (Ax)^T(Ax) = x^T A^T Ax = x^T(A^T A)x.$$

Solution. The quantity $\|Ax\|^2$ is maximized at the same x that maximizes $\|Ax\|$, and $\|Ax\|^2$ is easier to study. Observe that

$$\|Ax\|^2 = (Ax)^T(Ax) = x^T A^T Ax = x^T(A^T A)x.$$

Also, $A^T A$ is a symmetric matrix, since

$$(A^T A)^T = A^T A^{TT} = A^T A.$$

So the problem now is

$$\begin{array}{ll} \text{to maximize:} & x^T(A^TA)x \\ \text{subject to:} & \|x\| = 1. \end{array}$$

So the problem now is

$$\begin{array}{ll} \text{to maximize:} & x^T(A^TA)x \\ \text{subject to:} & \|x\| = 1. \end{array}$$

We know that the maximum value is the greatest eigenvalue λ_1 of A^TA.

So the problem now is

$$\begin{array}{ll} \text{to maximize:} & x^T(A^TA)x \\ \text{subject to:} & \|x\| = 1. \end{array}$$

We know that the maximum value is the greatest eigenvalue λ_1 of A^TA. Also, the maximum value is attained at a unit eigenvector of A^TA corresponding to λ_1.

For the matrix A in this example,

$$A^T A = \begin{pmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{pmatrix} \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

For the matrix A in this example,

$$A^TA = \begin{pmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{pmatrix} \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

The eigenvalues of A^TA are

$$\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0.$$

For the matrix A in this example,

$$A^T A = \begin{pmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{pmatrix} \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

The eigenvalues of $A^T A$ are

$$\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0.$$

Corresponding unit eigenvectors are, respectively,

$$v_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, v_2 = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}, v_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}.$$

The maximum value of $\|Ax\|^2$ is 360, attained when x is the unit vector v_1.

The maximum value of $\|Ax\|^2$ is 360, attained when x is the unit vector v_1. The vector Av_1 is a point on the ellipse in the above figure farthest from the origin,

The maximum value of $\|Ax\|^2$ is 360, attained when x is the unit vector v_1. The vector Av_1 is a point on the ellipse in the above figure farthest from the origin, namely,

$$Av_1 = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} 18 \\ 6 \end{pmatrix}.$$

The maximum value of $\|Ax\|^2$ is 360, attained when x is the unit vector v_1. The vector Av_1 is a point on the ellipse in the above figure farthest from the origin, namely,

$$Av_1 = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} 18 \\ 6 \end{pmatrix}.$$

Therefore, for $\|x\| = 1$, the maximum value of $\|Ax\|$ is

$$\|Av_1\| = \sqrt{360} = 6\sqrt{10}.$$

The last two examples suggest that

- The effect of A on the unit sphere in $\mathbb{R}^3$ is related to the quadratic form $x^T(A^TA)x$.

The last two examples suggest that

- The effect of A on the unit sphere in $\mathbb{R}^3$ is related to the quadratic form $x^T(A^TA)x$.
- To decompose an $m \times n$, it is worthwhile to study A^TA or AA^T. Both are symmetric square matrices and can be orthogonally diagonalized, as we know.

Singular Values of an $m \times n$ Matrix

Since matrix products of the form $A^T A$ will play an important role in our further work, we now give some basic properties of $A^T A$.

Since matrix products of the form A^TA will play an important role in our further work, we now give some basic properties of A^TA.

Theorem.

Let A be an $m \times n$ matrix. Then

1. A^TA is a square matrix.
2. A^TA is symmetric and so it is orthogonally diagonalizable.
3. All the eigenvalues of A^TA are non-negative.

Proof. Let A be an $m \times n$ matrix. Then

1. We observe that

$$(A^T)_{n\times m}(A)_{m\times n} = (A^T A)_{n\times n}.$$

Proof. Let A be an $m \times n$ matrix. Then

1. We observe that

$$(A^T)_{n\times m}(A)_{m\times n} = (A^T A)_{n\times n}.$$

This shows that $A^T A$ is a square matrix of size $n \times n$.

Proof. Let A be an $m \times n$ matrix. Then

1. We observe that

$$(A^T)_{n\times m}(A)_{m\times n} = (A^TA)_{n\times n}.$$

This shows that A^TA is a square matrix of size $n \times n$.

2. We observe that

$$(A^TA)^T = A^T(A^T)^T = A^TA.$$

Proof. Let A be an $m \times n$ matrix. Then

1. We observe that

$$(A^T)_{n\times m}(A)_{m\times n} = (A^TA)_{n\times n}.$$

This shows that A^TA is a square matrix of size $n \times n$.

2. We observe that

$$(A^TA)^T = A^T(A^T)^T = A^TA.$$

This shows that A^TA is symmetric. Then by the spectral decomposition theorem, A^TA is orthogonally diagonalizable.

3. By (2), the matrix $A^T A$ is symmetric and can be orthogonally diagonalized.

3. By (2), the matrix A^TA is symmetric and can be orthogonally diagonalized. Let $\{v_1, \cdots, v_n\}$ be eigenvectors of A^TA, and let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues of A^TA.

3. By (2), the matrix A^TA is symmetric and can be orthogonally diagonalized. Let $\{v_1, \cdots, v_n\}$ be eigenvectors of A^TA, and let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \leq i \leq n$,

$$\|Av_i\|^2 = (Av_i)^T Av_i = v_i^T A^T Av_i$$

3. By (2), the matrix A^TA is symmetric and can be orthogonally diagonalized. Let $\{v_1, \cdots, v_n\}$ be eigenvectors of A^TA, and let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \leq i \leq n$,

$$\begin{aligned}\|Av_i\|^2 &= (Av_i)^T Av_i = v_i^T A^T Av_i \\ &= v_i^T(\lambda_i v_i) = \lambda_i(v_i^T v_i)\end{aligned}$$

3. By (2), the matrix A^TA is symmetric and can be orthogonally diagonalized. Let $\{v_1, \cdots, v_n\}$ be eigenvectors of A^TA, and let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \le i \le n$,

$$\begin{aligned}\|Av_i\|^2 &= (Av_i)^T Av_i = v_i^T A^T Av_i \\ &= v_i^T(\lambda_i v_i) = \lambda_i(v_i^T v_i) \\ &= \lambda_i\end{aligned}$$

3. By (2), the matrix A^TA is symmetric and can be orthogonally diagonalized. Let $\{v_1, \cdots, v_n\}$ be eigenvectors of A^TA, and let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \leq i \leq n$,

$$\begin{aligned} \|Av_i\|^2 &= (Av_i)^T Av_i = v_i^T A^T Av_i \\ &= v_i^T(\lambda_i v_i) = \lambda_i(v_i^T v_i) \\ &= \lambda_i \end{aligned}$$

So the eigenvalues of A^TA are all nonnegative.

We proved that the eigenvalues of $A^T A$ are all nonnegative. It therefore makes sense to take (positive) square roots of these eigenvalues.

We proved that the eigenvalues of $A^T A$ are all nonnegative. It therefore makes sense to take (positive) square roots of these eigenvalues. We give them special name.

Singular values

The square roots of the eigenvalues of $A^T A$ are called the **singular values** of A.

If $\lambda_1, ..., \lambda_n$ be the eigenvalues of $A^T A$, then the corresponding singular values of A are denoted by

$$\sigma_1, \sigma_2, ..., \sigma_n,$$

If $\lambda_1, ..., \lambda_n$ be the eigenvalues of $A^T A$, then the corresponding singular values of A are denoted by

$$\sigma_1, \sigma_2, ..., \sigma_n,$$

and

$$\sigma_i = \sqrt{\lambda_i} \quad 1 \leq i \leq n.$$

If $\lambda_1, ..., \lambda_n$ be the eigenvalues of $A^T A$, then the corresponding singular values of A are denoted by

$$\sigma_1, \sigma_2, ..., \sigma_n,$$

and

$$\sigma_i = \sqrt{\lambda_i} \quad 1 \leq i \leq n.$$

Problem

The singular values of A are the lengths of the vectors $Av_1, ..., Av_n$.

Example.

Find the singular values of the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Example.

Find the singular values of the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Solution. Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Example.

Find the singular values of the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Solution. Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The first step is to find the eigenvalues of the matrix $A^T A$.

We have

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

We have

$$A^T A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The characteristic polynomial of $A^T A$ is

$$\lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

We have

$$A^T A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The characteristic polynomial of $A^T A$ is

$$\lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

so the eigenvalues of $A^T A$ are $\lambda_1 = 3$ and $\lambda_2 = 1$

We have

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

The characteristic polynomial of A^TA is

$$\lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

so the eigenvalues of A^TA are $\lambda_1 = 3$ and $\lambda_2 = 1$ and the singular values of A in order of decreasing size are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3},$$
$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1.$$

Theorem

If A is an $m \times n$ matrix, then $A^T A$ and AA^T have the same nonzero eigenvalues.

Theorem

If A is an $m \times n$ matrix, then $A^T A$ and AA^T have the same nonzero eigenvalues.

Proof. Suppose A is an $m \times n$ matrix, and suppose that λ is a nonzero eigenvalue of $A^T A$.

Then there exists a nonzero vector $x \in \mathbb{R}^n$ such that
$$(A^T A)x = \lambda x.$$

Then there exists a nonzero vector $x \in \mathbb{R}^n$ such that

$$(A^T A)x = \lambda x.$$

Multiplying both sides of this equation by A:

$$A(A^T A)x = A\lambda x$$

$$\Rightarrow (AA^T)(Ax) = \lambda(Ax).$$

Then there exists a nonzero vector $x \in \mathbb{R}^n$ such that
$$(A^T A)x = \lambda x.$$
Multiplying both sides of this equation by A:
$$A(A^T A)x = A\lambda x$$
$$\Rightarrow (AA^T)(Ax) = \lambda(Ax).$$
Since $\lambda \neq 0$ and $x \neq 0$, $\lambda x \neq 0$, and thus, $(A^T A)x \neq 0$; thus $A^T(Ax) \neq 0$, implying that $Ax \neq 0$.

Then there exists a nonzero vector $x \in \mathbb{R}^n$ such that

$$(A^T A)x = \lambda x.$$

Multiplying both sides of this equation by A:

$$A(A^T A)x = A\lambda x$$

$$\Rightarrow (AA^T)(Ax) = \lambda(Ax).$$

Since $\lambda \neq 0$ and $x \neq 0$, $\lambda x \neq 0$, and thus, $(A^T A)x \neq 0$; thus $A^T(Ax) \neq 0$, implying that $Ax \neq 0$. Therefore Ax is an eigenvector of AA^T corresponding to eigenvalue λ.

Then there exists a nonzero vector $x \in \mathbb{R}^n$ such that

$$(A^T A)x = \lambda x.$$

Multiplying both sides of this equation by A:

$$A(A^T A)x = A\lambda x$$

$$\Rightarrow (AA^T)(Ax) = \lambda(Ax).$$

Since $\lambda \neq 0$ and $x \neq 0$, $\lambda x \neq 0$, and thus, $(A^T A)x \neq 0$; thus $A^T(Ax) \neq 0$, implying that $Ax \neq 0$. Therefore Ax is an eigenvector of AA^T corresponding to eigenvalue λ. An analogous argument can be used to show that every nonzero eigenvalue of AA^T is an eigenvalue of $A^T A$, thus completing the proof. $\square$

- For a symmetric matrix, the number of its nonzero eigenvalues is its rank.

- For a symmetric matrix, the number of its nonzero eigenvalues is its rank.
- To obtain a similar result for the rank of an $m \times n$ matrix A, not necessarily square, we arrange the eigenvalues of $A^T A$ in decreasing order:

$$\lambda_1 \geq \cdots \geq \lambda_n \geq 0.$$

- For a symmetric matrix, the number of its nonzero eigenvalues is its rank.
- To obtain a similar result for the rank of an $m \times n$ matrix A, not necessarily square, we arrange the eigenvalues of $A^T A$ in decreasing order:
$$\lambda_1 \geq \cdots \geq \lambda_n \geq 0.$$
Thus, by definition, the corresponding singular values of A follow the same order:
$$\sigma_1 \geq \cdots \geq \sigma_n.$$

The decomposition of A involves an $m \times n$ "diagonal" matrix Σ of the form

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \\ \leftarrow m - r \text{ rows} \end{array}$$

$$\uparrow \; n - r \text{ columns}$$

where D is an $r \times r$ diagonal matrix for some r not exceeding the smaller of m and n. (If r equals m or n or both, some or all of the zero matrices do not appear.)

Theorem

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix Σ as in the above for which the diagonal entries in D are the first r singular values of A,

$$\sigma_1 \geq \cdots \geq \sigma_r > 0$$

and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

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

Any factorization $A = U\Sigma V^T$, with U and V orthogonal, Σ as in the above, and positive diagonal entries in D, is called a **singular value decomposition** (or SVD) of A.

Any factorization $A = U\Sigma V^T$, with U and V orthogonal, Σ as in the above, and positive diagonal entries in D, is called a **singular value decomposition** (or SVD) of A.
The matrices U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A.

Any factorization $A = U\Sigma V^T$, with U and V orthogonal, Σ as in the above, and positive diagonal entries in D, is called a **singular value decomposition** (or SVD) of A.

The matrices U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A.

The columns of U in such a decomposition are called **left singular vectors** of A, and the columns of V are called **right singular vectors** of A.

Geometry

The theorem allows to decompose the action of A on a given input vector as a three-step process. To get Ax, where x in $\mathbb{R}^n$, we first form $\tilde{x} := V^T x$ in $\mathbb{R}^n$. Since V is an orthogonal matrix, V^T is also orthogonal, and $\tilde{x}$ is just a rotated version of x, which still lies in the input space. Then we act on the rotated vector $\tilde{x}$ by scaling its elements.

Precisely, the first r elements of $\tilde{x}$ are scaled by the singular values $\sigma_1, \ldots, \sigma_r$; the remaining $n - r$ elements are set to zero. This step results in a new vector $\tilde{y}$ which now belongs to the output space $\mathbb{R}^m$. The final step consists in rotating the vector $\tilde{y}$ by the orthogonal matrix U, which results in $y = U\tilde{y} = Ax$.

Precisely, the first r elements of $\tilde{x}$ are scaled by the singular values $\sigma_1, \ldots, \sigma_r$; the remaining $n-r$ elements are set to zero. This step results in a new vector $\tilde{y}$ which now belongs to the output space $\mathbb{R}^m$. The final step consists in rotating the vector $\tilde{y}$ by the orthogonal matrix U, which results in $y = U\tilde{y} = Ax$.

Corollary (EVD in terms of outer poducts).
The equation $A = U\Sigma V^T$ can be written as

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T.$$

The following summary of the proof also provides an algorithm to obtain an SVD of an $m \times n$ real matrix A. Note that by construction U and V are not unique, especially if an eigenvalue of $A_T A$ repeats. However, the matrix Σ is unique.

Step 1. Find the eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n \geq 0.$$

of $A^T A$. Put $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \ldots, r$ (r the largest index with $\lambda_j > 0$). Take

$$D = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$$

and extend D to an $m \times n$ matrix Σ with the block D at its upper left corner, zeros elsewhere.

Step 2. Find unit eigenvectors $v_1, \dots v_n$ corresponding ing to λ_j. Take V to be the $n \times n$ (orthonormal) matrix with v_j as its j-th column.

Step 2. Find unit eigenvectors $v_1, \dots v_n$ corresponding ing to λ_j. Take V to be the $n \times n$ (orthonormal) matrix with v_j as its j-th column.
Step 3. For $j = 1, \dots, r$, put $u_j = \frac{1}{\sigma_j} A v_j$. Then $\{u_1, \dots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$. Extend it to an orthonormal basis $\{u_1, \dots, u_m\}$ of $\mathbb{R}^m$. Take U to be the $m \times m$ orthogonal matrix with u_j as its j-th column.

Step 2. Find unit eigenvectors $v_1, \ldots v_n$ corresponding ing to λ_j. Take V to be the $n \times n$ (orthonormal) matrix with v_j as its j-th column.
Step 3. For $j = 1, \ldots, r$, put $u_j = \frac{1}{\sigma_j} A v_j$. Then $\{u_1, \ldots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$. Extend it to an orthonormal basis $\{u_1, \ldots, u_m\}$ of $\mathbb{R}^m$. Take U to be the $m \times m$ orthogonal matrix with u_j as its j-th column.
Step 4. Check that

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

Theorem

Let A be an $m \times n$ matrix and let $B = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis of n consisting of eigenvectors of $A^T A$, with corresponding eigenvectors $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then Av_i is orthogonal to Av_j for $i \neq j$.

Since B is an orthonormal basis of $\mathbb{R}^n$, if $i \neq j$, then

$$(Av_i) \cdot (Av_j) = (Av_i)^T (Av_j)$$

Since B is an orthonormal basis of $\mathbb{R}^n$, if $i \neq j$, then

$$\begin{aligned}(Av_i) \cdot (Av_j) &= (Av_i)^T (Av_j) \\ &= v_i^T (A^T A) v_j\end{aligned}$$

Since B is an orthonormal basis of $\mathbb{R}^n$, if $i \neq j$, then

$$\begin{aligned}(Av_i) \cdot (Av_j) &= (Av_i)^T (Av_j) \\ &= v_i^T (A^T A) v_j \\ &= v_i^T \lambda_j v_i\end{aligned}$$

Since B is an orthonormal basis of $\mathbb{R}^n$, if $i \neq j$, then

$$\begin{aligned}(Av_i) \cdot (Av_j) &= (Av_i)^T (Av_j) \\ &= v_i^T (A^T A) v_j \\ &= v_i^T \lambda_j v_i \\ &= \lambda_j v_i^T v_j\end{aligned}$$

Since B is an orthonormal basis of $\mathbb{R}^n$, if $i \neq j$, then

$$\begin{aligned}
(Av_i) \cdot (Av_j) &= (Av_i)^T (Av_j) \\
&= v_i^T (A^T A) v_j \\
&= v_i^T \lambda_j v_i \\
&= \lambda_j v_i^T v_j \\
&= 0.
\end{aligned}$$

$\square$

Examples

Example (Non-symmetric matrix):
Find the SVD of the matrix:

$$A = \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix}.$$

Solution.

Step 1. We have

$$A^T A = \begin{pmatrix} 2 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}.$$

Solution.

Step 1. We have

$$A^T A = \begin{pmatrix} 2 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}.$$

To find the eigenvalues of $A^T A$, we have

$$\begin{vmatrix} 5-\lambda & -3 \\ -3 & 5-\lambda \end{vmatrix} = (5-\lambda)^2 - 9.$$

Solution.

Step 1. We have

$$A^T A = \begin{pmatrix} 2 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}.$$

To find the eigenvalues of $A^T A$, we have

$$\begin{vmatrix} 5-\lambda & -3 \\ -3 & 5-\lambda \end{vmatrix} = (5-\lambda)^2 - 9.$$

Clearly, the eigenvalues of $A^T A$ are 8 and 2. So its singular values are $\sigma_1 = 2\sqrt{2}$ and$\sigma_2 = \sqrt{2}$, and

$$\Sigma = \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}.$$

Step. 2 We find the eigenvectors corresponding to eigenvalues λ_1 andλ_2.

Step. 2 We find the eigenvectors corresponding to eigenvalues λ_1 andλ_2. For $\lambda_1 = 8$, we have

$$\begin{pmatrix} -3 & -3 \\ -3 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

Step. 2 We find the eigenvectors corresponding to eigenvalues λ_1 andλ_2. For $\lambda_1 = 8$, we have

$$\begin{pmatrix} -3 & -3 \\ -3 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 + x_2 = 0.$$

Step. 2 We find the eigenvectors corresponding to eigenvalues λ_1 andλ_2. For $\lambda_1 = 8$, we have

$$\begin{pmatrix} -3 & -3 \\ -3 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 + x_2 = 0.$$

So,

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

For $\lambda_2 = 2$, we have

$$\begin{pmatrix} 3 & -3 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

For $\lambda_2 = 2$, we have

$$\begin{pmatrix} 3 & -3 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 - x_2 = 0.$$

For $\lambda_2 = 2$, we have

$$\begin{pmatrix} 3 & -3 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 - x_2 = 0.$$

So,

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For $\lambda_2 = 2$, we have

$$\begin{pmatrix} 3 & -3 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 - x_2 = 0.$$

So,

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Thus,

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Step 3. We now find U. We have

$$u_1 = \frac{1}{\sigma_1} A v_1$$

Step 3. We now find U. We have

$$u_1 = \frac{1}{\sigma_1} A v_1$$
$$= \frac{1}{\sqrt{8}} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Step 3. We now find U. We have

$$\begin{aligned} u_1 &= \frac{1}{\sigma_1} A v_1 \\ &= \frac{1}{\sqrt{8}} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 \\ 0 \end{pmatrix}. \end{aligned}$$

Similarly, we have

$$u_2 = \frac{1}{\sigma_2} A v_2$$

Similarly, we have

$$\begin{aligned} u_2 &= \frac{1}{\sigma_2} A v_2 \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{aligned}$$

Similarly, we have

$$\begin{aligned} u_2 &= \frac{1}{\sigma_2} A v_2 \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Similarly, we have

$$
\begin{aligned}
u_2 &= \frac{1}{\sigma_2} A v_2 \\
&= \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 1 \end{pmatrix}.
\end{aligned}
$$

Thus,

$$
U = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.
$$

Step 4. Check if we got the SVD of A:

$$U\Sigma V^T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

Step 4. Check if we got the SVD of A:

$$U\Sigma V^T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix}$$

Step 4. Check if we got the SVD of A:

$$
\begin{aligned}
U\Sigma V^T &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} \\
&= A.
\end{aligned}
$$

Example (Rectangular matrix):
Find the SVD of the matrix:

$$A = \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}.$$

Solution.
Step 1. We have

$$A^T A = \begin{pmatrix} 1 & -2 & 2 \\ -1 & 2 & -2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} = 9 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $A^T A = 9B$, the eigenvalues of $A^T A$ are 9 times those of B.

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $A^TA = 9B$, the eigenvalues of A^TA are 9 times those of B. To find the eigenvalues of B, we have

$$\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 1.$$

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $A^T A = 9B$, the eigenvalues of $A^T A$ are 9 times those of B. To find the eigenvalues of B, we have

$$\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 1.$$

Clearly, the eigenvalues of B are 2 and 0.

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $A^T A = 9B$, the eigenvalues of $A^T A$ are 9 times those of B. To find the eigenvalues of B, we have

$$\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 1.$$

Clearly, the eigenvalues of B are 2 and 0. So, the eigenvalues of $A^T A$ are $\lambda_1 = 18$ and $\lambda_2 = 0$.

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $A^T A = 9B$, the eigenvalues of $A^T A$ are 9 times those of B. To find the eigenvalues of B, we have

$$\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 1.$$

Clearly, the eigenvalues of B are 2 and 0. So, the eigenvalues of $A^T A$ are $\lambda_1 = 18$ and $\lambda_2 = 0$. Thus, its singular values are $\sigma_1 = 3\sqrt{2}$ and$\sigma_2 = 0$,

Put $B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $A^T A = 9B$, the eigenvalues of $A^T A$ are 9 times those of B. To find the eigenvalues of B, we have

$$\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 1.$$

Clearly, the eigenvalues of B are 2 and 0. So, the eigenvalues of $A^T A$ are $\lambda_1 = 18$ and $\lambda_2 = 0$. Thus, its singular values are $\sigma_1 = 3\sqrt{2}$ and$\sigma_2 = 0$, and

$$\Sigma = \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Step. 2 Since $A^TA = 9B$, the eigenvectors of A^TA are those of B.

Step. 2 Since $A^TA = 9B$, the eigenvectors of A^TA are those of B. We find the eigenvectors corresponding to eigenvalues 2 and0 of B.

Step. 2 Since $A^T A = 9B$, the eigenvectors of $A^T A$ are those of B. We find the eigenvectors corresponding to eigenvalues 2 and0 of B. For $\lambda_1 = 2$, we have

$$\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

Step. 2 Since $A^TA = 9B$, the eigenvectors of A^TA are those of B. We find the eigenvectors corresponding to eigenvalues 2 and0 of B. For $\lambda_1 = 2$, we have

$$\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 + x_2 = 0.$$

Step. 2 Since $A^TA = 9B$, the eigenvectors of A^TA are those of B. We find the eigenvectors corresponding to eigenvalues 2 and0 of B. For $\lambda_1 = 2$, we have

$$\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 + x_2 = 0.$$

So,

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

For $\lambda_2 = 0$, we have

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

For $\lambda_2 = 0$, we have

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 - x_2 = 0.$$

For $\lambda_2 = 0$, we have

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 - x_2 = 0.$$

So,

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For $\lambda_2 = 0$, we have

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

This gives

$$x_1 - x_2 = 0.$$

So,

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Thus,

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Step 3. We now find U. We have

$$
\begin{aligned}
u_1 &= \frac{1}{\sigma_1} A v_1 \\
&= \frac{1}{\sqrt{18}} \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}
\end{aligned}
$$

Step 3. We now find U. We have

$$u_1 = \frac{1}{\sigma_1} A v_1$$

$$= \frac{1}{\sqrt{18}} \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$= \frac{1}{3} \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}.$$

We have only one nonzero eigenvalue.

We have only one nonzero eigenvalue. So, we have to extend $\{u_1\}$ to an orthonormal basis of $\mathbb{R}^3$.

We have only one nonzero eigenvalue. So, we have to extend $\{u_1\}$ to an orthonormal basis of $\mathbb{R}^3$. For it we pick any orthonormal basis $\{u_2, u_3\}$ of the orthogonal complement of u_1, i.e. the solution space of the equation

$$u_1 \cdot x = 0$$

$$\Rightarrow \frac{1}{3}\begin{pmatrix} 1 & -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

$$\Rightarrow x - 2y + 2z = 0.$$

The vectors satisfying the last equation are

$$w_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

The vectors satisfying the last equation are

$$w_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

We now use Gram-Schmidt orthonormalisation method.

The vectors satisfying the last equation are

$$w_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

We now use Gram-Schmidt orthonormalisation method. Put

$$u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

For u_3, we have

$$w_3 - (w_3 \cdot u_2)u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 4/5 \\ 1 \end{pmatrix}.$$

For u_3, we have

$$w_3 - (w_3 \cdot u_2)u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 4/5 \\ 1 \end{pmatrix}.$$

Now, put

$$u_3 = \frac{\sqrt{45}}{5}\begin{pmatrix} -2 \\ 4 \\ 5 \end{pmatrix}.$$

For u_3, we have

$$w_3 - (w_3 \cdot u_2)u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 4/5 \\ 1 \end{pmatrix}.$$

Now, put

$$u_3 = \frac{\sqrt{45}}{5}\begin{pmatrix} -2 \\ 4 \\ 5 \end{pmatrix}.$$

Thus,

$$U = (u_1 \ u_2 \ u_3).$$

For u_3, we have

$$w_3 - (w_3 \cdot u_2)u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 4/5 \\ 1 \end{pmatrix}.$$

Now, put

$$u_3 = \frac{\sqrt{45}}{5}\begin{pmatrix} -2 \\ 4 \\ 5 \end{pmatrix}.$$

Thus,

$$U = (u_1 \; u_2 \; u_3).$$

Step 4. Finally, check that $U\Sigma V^T = A$.

Four fundamental subspaces of a matrix

Suppose that A is an $m \times n$ matrix that maps vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$.

Suppose that A is an $m \times n$ matrix that maps vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$. The four fundamental subspaces of A, two in $\mathbb{R}^n$ and two in $\mathbb{R}^m$, are:

Column space,Row space, Null space, Left null space

Column space

Column space of A is the span of the columns of A. In symbols,

$$\mathrm{Col}(A) = \{Ax : x \in \mathbb{R}^n\}.$$

Column space

Column space of A is the span of the columns of A. In symbols,

$$\mathrm{Col}(A) = \{Ax : x \in \mathbb{R}^n\}.$$

Row space

Row space of A is the span of the rows of A. In symbols,

$$\mathrm{Row}(A) = \{A^T y : y \in \mathbb{R}^m\}$$

Null space

Null space of A is denoted by $\mathrm{Null}(A)$ and is defined by

$$\mathrm{Null}(A) = \{x \in \mathbb{R}^n : \; Ax = 0\}$$

Null space

Null space of A is denoted by $\mathrm{Null}(A)$ and is defined by

$$\mathrm{Null}(A) = \{x \in \mathbb{R}^n : \; Ax = 0\}$$

Left null space

Left null space of A, the set of all y for which $A^T y = 0$. In symbols,

$$\mathrm{Null}(A^T) = \{y \in \mathbb{R}^m : \; A^T y = 0\}.$$

Dimension and rank.

We know that the dimension of a subspace is the number of linearly independent vectors required to span that space.

Dimension and rank.

We know that the dimension of a subspace is the number of linearly independent vectors required to span that space. The **Fundamental Theorem of Linear Algebra** is

Fundamental Theorem of Linear Algebra

The dimension of the row space is equal to the dimension of the column space.

Dimension and rank.

We know that the dimension of a subspace is the number of linearly independent vectors required to span that space. The **Fundamental Theorem of Linear Algebra** is

Fundamental Theorem of Linear Algebra

The dimension of the row space is equal to the dimension of the column space.

In other words, the number of linearly independent rows is equal to the number of linearly independent columns.

Dimension and rank.

We know that the dimension of a subspace is the number of linearly independent vectors required to span that space. The **Fundamental Theorem of Linear Algebra** is

Fundamental Theorem of Linear Algebra

The dimension of the row space is equal to the dimension of the column space.

In other words, the number of linearly independent rows is equal to the number of linearly independent columns. Hence

$$\dim \operatorname{Row}(A) = \dim \operatorname{Col}(A).$$

The **rank of a matrix** is this number of linearly independent rows or columns.

The **rank of a matrix** is this number of linearly independent rows or columns.
The following theorem gives relationships among them.

Theorem

Let A be an $m \times n$ matrix. Then

1. $\mathrm{Null}(A) = \mathrm{Row}(A)^{\perp}$.
2. $\mathrm{Null}(A^T) = \mathrm{Col}(A)^{\perp}$.

Proof.

(a) Let A be an $m \times n$ matrix.

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \text{Null}(A)$.

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \text{Null}(A)$. Then $Ax = 0$.

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \text{Null}(A)$. Then $Ax = 0$. Now, let $A^T y \in \text{Row}(A)$.

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \mathrm{Null}(A)$. Then $Ax = 0$. Now, let $A^T y \in \mathrm{Row}(A)$. Then

$$(A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax) = 0.$$

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \text{Null}(A)$. Then $Ax = 0$. Now, let $A^T y \in \text{Row}(A)$. Then

$$(A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax) = 0.$$

Since $A^T y$ is an arbitrary vector in $\text{Row}(A)$,

$$x \perp \text{Row}(A) \Rightarrow x \in \text{Row}(A)^{\perp}.$$

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \mathrm{Null}(A)$. Then $Ax = 0$. Now, let $A^T y \in \mathrm{Row}(A)$. Then

$$(A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax) = 0.$$

Since $A^T y$ is an arbitrary vector in $\mathrm{Row}(A)$,

$$x \perp \mathrm{Row}(A) \Rightarrow x \in \mathrm{Row}(A)^{\perp}.$$

Moreover, $x \in \mathrm{Null}(A)$ is arbitrary.

Proof.

(a) Let A be an $m \times n$ matrix. Assume that $x \in \text{Null}(A)$. Then $Ax = 0$. Now, let $A^T y \in \text{Row}(A)$. Then

$$(A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax) = 0.$$

Since $A^T y$ is an arbitrary vector in $\text{Row}(A)$,

$$x \perp \text{Row}(A) \Rightarrow x \in \text{Row}(A)^{\perp}.$$

Moreover, $x \in \text{Null}(A)$ is arbitrary. Therefore,

$$\text{Null}(A) \subseteq \text{Row}(A)^{\perp}.$$

Proof...

On the other hand, let $x \in \text{Row}(A)^{\perp}$.

Proof...

On the other hand, let $x \in \text{Row}(A)^{\perp}$. Then for any $A^T y \in \text{Row}(A)$, we have

$$0 = (A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax).$$

Proof...

On the other hand, let $x \in \mathrm{Row}(A)^{\perp}$. Then for any $A^T y \in \mathrm{Row}(A)$, we have

$$0 = (A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T(Ax).$$

Since $A^T y \in \mathrm{Row}(A)$ is arbitrary, we can choose $y \neq 0$ and so, $y^T \neq 0$.

Proof...

On the other hand, let $x \in \text{Row}(A)^{\perp}$. Then for any $A^T y \in \text{Row}(A)$, we have

$$0 = (A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax).$$

Since $A^T y \in \text{Row}(A)$ is arbitrary, we can choose $y \neq 0$ and so, $y^T \neq 0$. Therefore, we must have

$$Ax = 0.$$

That means, $x \in \text{Null}(A)$.

Proof...

On the other hand, let $x \in \mathrm{Row}(A)^{\perp}$. Then for any $A^T y \in \mathrm{Row}(A)$, we have

$$0 = (A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax).$$

Since $A^T y \in \mathrm{Row}(A)$ is arbitrary, we can choose $y \neq 0$ and so, $y^T \neq 0$. Therefore, we must have

$$Ax = 0.$$

That means, $x \in \mathrm{Null}(A)$.
Also, $A^T y \in \mathrm{Row}(A)$ is arbitrary.

Proof...

On the other hand, let $x \in \mathrm{Row}(A)^{\perp}$. Then for any $A^T y \in \mathrm{Row}(A)$, we have

$$0 = (A^T y) \cdot x = (A^T y)^T x = (y^T A) x = y^T (Ax).$$

Since $A^T y \in \mathrm{Row}(A)$ is arbitrary, we can choose $y \neq 0$ and so, $y^T \neq 0$. Therefore, we must have

$$Ax = 0.$$

That means, $x \in \mathrm{Null}(A)$.
Also, $A^T y \in \mathrm{Row}(A)$ is arbitrary. Therefore,

$$\mathrm{Row}(A)^{\perp} \subseteq \mathrm{Null}(A).$$

Proof...

On the other hand, let $x \in \text{Row}(A)^{\perp}$. Then for any $A^T y \in \text{Row}(A)$, we have

$$0 = (A^T y) \cdot x = (A^T y)^T x = (y^T A)x = y^T (Ax).$$

Since $A^T y \in \text{Row}(A)$ is arbitrary, we can choose $y \neq 0$ and so, $y^T \neq 0$. Therefore, we must have

$$Ax = 0.$$

That means, $x \in \text{Null}(A)$.
Also, $A^T y \in \text{Row}(A)$ is arbitrary. Therefore,

$$\text{Row}(A)^{\perp} \subseteq \text{Null}(A).$$

Hence

$$\text{Null}(A) = \text{Row}(A)^{\perp}.$$

(b) Substitute A^T for A in part (a) and note that $\text{Row}(A^T) = \text{Col}(A)$.

(b) Substitute A^T for A in part (a) and note that $\text{Row}(A^T) = \text{Col}(A)$.

Theorem

Let A be an $m \times n$ matrix $A = U\Sigma V^T$ be any SVD for A where U and V are orthogonal of size $m \times m$ and $n \times n$ respectively, and rank $A = r$. Also, let

$$\begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}.$$

where $D = \underline{\text{diag}}(\lambda_1, ..., \lambda_r)$ with each $\lambda_i > 0$.

If $U = \{u_1 \ldots u_r \ldots u_m\}$ and $V = \{v_1 \ldots v_r \ldots v_n\}$ are orthonormal bases for $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, then

1. $\{u_1 \ldots u_r\}$ is an orthonormal basis of Col A
2. $\{u_{r+1} \ldots u_m\}$ is an orthonormal basis of Null A^T
3. $\{v_{r+1} \ldots v_n\}$ is an orthonormal basis of Null A
4. $\{v_1 \ldots v_r\}$ is an orthonormal basis of Row A

Principal Component Analysis

Qustion: What does PCA compute and then what is PCA doing with the knowledge of SVD ?

Principal Component Analysis

Qustion: What does PCA compute and then what is PCA doing with the knowledge of SVD ?
Consider a real-valued data matrix X that is $n \times p$ where n is the number of samples and p is the number of features. Its SVD is

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

Now consider the covariance matrix of X, namely $X^T X$, in the context of the SVD decomposition:

Now consider the covariance matrix of X, namely X^TX, in the context of the SVD decomposition:

$$X^TX = (U\Sigma V^T)^T(U\Sigma V^T)$$

Now consider the covariance matrix of X, namely $X^T X$, in the context of the SVD decomposition:

$$\begin{aligned} X^T X &= (U\Sigma V^T)^T (U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \end{aligned}$$

Now consider the covariance matrix of X, namely X^TX, in the context of the SVD decomposition:

$$\begin{aligned} X^TX &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T\Sigma V^T \end{aligned}$$

Now consider the covariance matrix of X, namely X^TX, in the context of the SVD decomposition:

$$\begin{aligned} X^TX &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T\Sigma V^T \\ &= V\Sigma^2 V^T \end{aligned}$$

Now consider the covariance matrix of X, namely $X^T X$, in the context of the SVD decomposition:

$$\begin{aligned} X^T X &= (U\Sigma V^T)^T (U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T \Sigma V^T \\ &= V\Sigma^2 V^T \end{aligned}$$

We already know that V and V^T are just rotation matrices, while U disappears because $U^T U = I$.

Now consider the covariance matrix of X, namely X^TX, in the context of the SVD decomposition:

$$
\begin{aligned}
X^TX &= (U\Sigma V^T)^T(U\Sigma V^T) \\
&= V\Sigma^T U^T U\Sigma V^T \\
&= V\Sigma^T\Sigma V^T \\
&= V\Sigma^2 V^T
\end{aligned}
$$

We already know that V and V^T are just rotation matrices, while U disappears because $U^TU = I$. And even without knowing about eigenvectors or eigenvalues, we can see what PCA is doing by understanding SVD: it is diagonalizing the covariance matrix of X.

Now consider the covariance matrix of X, namely $X^T X$, in the context of the SVD decomposition:

$$\begin{aligned} X^T X &= (U\Sigma V^T)^T (U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T \Sigma V^T \\ &= V\Sigma^2 V^T \end{aligned}$$

We already know that V and V^T are just rotation matrices, while U disappears because $U^T U = I$. And even without knowing about eigenvectors or eigenvalues, we can see what PCA is doing by understanding SVD: it is diagonalizing the covariance matrix of X. So PCA is finding the major axes along which our data varies.