Applied Linear Algebra

OTTO BRETSCHER

http://www.prenhall.com/bretscher

Chapter 8
Symmetric Matrices and Quadratic Forms

Chia-Hui Chang
Email: chia@csie.ncu.edu.tw
National Central University, Taiwan

8.1 SYMMETRIC MATRICES

In chapter 7, we are concerned with when is a given square matrix A diagonalizable? That is, when is there an `eigenbasis` for A?

In geometry, we prefer to work with `orthnomal bases`, which raises the question:

For which matrices is there an *orthonormal eigenbasis*?

Example 1 If A is orthogonally diagonalizable, what is the relationship between A^T and A?

Solution We have

$$S^{-1}AS = D$$

or

$$A = SDS^{-1} = SDS^T$$

for an orthogonal matrix S and a diagonal D. Then

$$A^T = (SDS^T)^T = SD^TS^T = SDS^T = A.$$

We find that A is symmetric.

Fact 8.1.1 Spectral theorem
A matrix A is orthogonally diagonalizable if and only if A is `symmetric` (i.e., $A^T = A$).

The set of eigenvalues of a matrix is called the spectrum of A, and the following description of the eigenvalues is called a spectral theorem.

THEOREM

The Spectral Theorem For A Symmetric Matrix

- A has n real eigenvalues, counting mutiplicities. (Fact 8.1.3)

- The dimension of the eigenspace for each eigenvalue λ equals the algebraic multiplicity of λ.

- The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal. (Fact 8.1.2)

- A is orthogonally diagonalizable. (Fact 8.1.1)

Example 2 For the symmetric matrix $A = \begin{bmatrix} 4 & 2 \\ 2 & 7 \end{bmatrix}$, find an orthogonal S such that $S^{-1}AS$ is diagonal.

Solution See Figure 1.

$$E_3 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, E_8 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

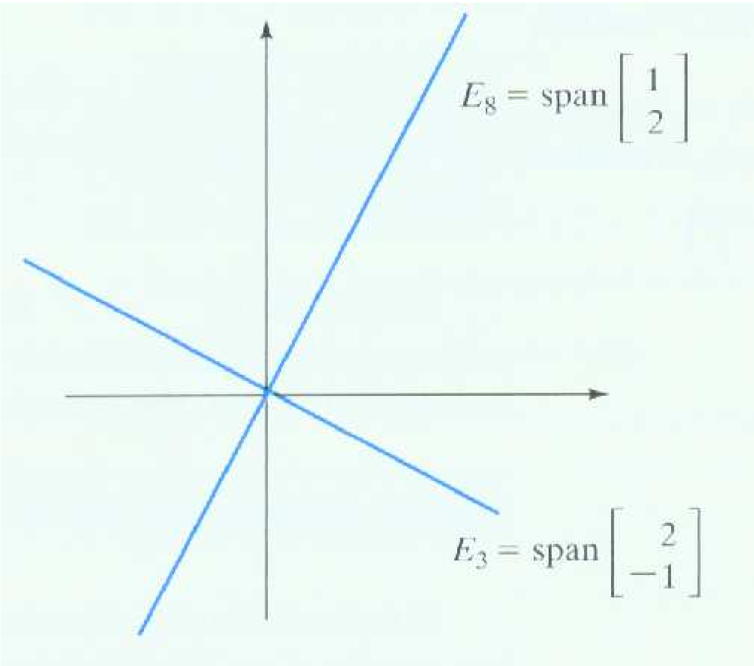


Figure 1

Note that the eigenspaces E_3 and E_8 are perpendicular. (This is no coincidence.) Therefore, we can find an orthonormal eigenbasis simply by dividing the given eigenvectors by their lengths:

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

Define

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

then $S^{-1}AS = \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix}$

Fact 8.1.2 Consider a symmetric matrix A. If $\vec{v}_1$ and $\vec{v}_2$ are eigenvectors of A with distinct eigenvalues λ_1 and λ_2, then $\vec{v}_1 \cdot \vec{v}_2 = 0$; that is, $\vec{v}_2$ is orthogonal to $\vec{v}_1$.

Proof We compute the product $\vec{v}_1^T A \vec{v}_2$ in two ways:

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

- $\vec{v}_1^T A \vec{v}_2 = \vec{v}_1^T A^T \vec{v}_2 = (A\vec{v}_1)^T \vec{v}_2 = (\lambda_1 \vec{v}_1)^T \vec{v}_2 = \lambda_1 (\vec{v}_1 \cdot \vec{v}_2)$

Comparing the results, we find

$$\lambda_1 (\vec{v}_1 \cdot \vec{v}_2) = \lambda_2 (\vec{v}_1 \cdot \vec{v}_2)$$

or

$$(\lambda_1 - \lambda_2)(\vec{v}_1 \cdot \vec{v}_2) = 0$$

Since $\lambda_1 \neq \lambda_2$, $\vec{v}_1 \cdot \vec{v}_2$ must be zero.

Fact 8.1.3 A symmetric $n \times n$ matrix A has n real eigenvalues if they are counted with their algebraic multiplicites.

Proof of 8.1.3 For those who have studied Section 7.5. Consider two complex conjugate eigenvalues $p \pm iq$ of A with corresponding eigenvectors $\vec{v} \pm i\vec{w}$. Compute the product

$$(\vec{v} + i\vec{w})^T A(\vec{v} - i\vec{w})$$

in two different ways:

$$(\vec{v} + i\vec{w})^T A(\vec{v} - i\vec{w}) = (\vec{v} + i\vec{w})^T (p - iq)(\vec{v} - i\vec{w})$$

$$= (p - iq)(\|\vec{v}\|^2 + \|\vec{w}\|^2)$$

$$(\vec{v} + i\vec{w})^T A(\vec{v} - i\vec{w}) = (A(\vec{v} + i\vec{w}))^T (\vec{v} - i\vec{w})$$

$$= (p+iq)(\vec{v}+i\vec{w})^T(\vec{v}-i\vec{w}) = (p+iq)(\|\vec{v}\|^2+\|\vec{w}\|^2)$$

Comparing the results, we find that $p + iq = p - iq$, so $q = 0$, as claimed.

Proof of 8.1.1 Even more technical.

Example 3 For the symmetric matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

find an orthogonal S such that $S^{-1}AS$ is diagonal.

Solution

The eigenvalues are 0 and 3, with

$E_0 = span\left(\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right)$ and $E_3 = span\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Note that the two eigenspaces are indeed perpendicular to one another (See Figure 2, 3).

We can construct an orthonormal eigenbasis for A by picking an orthonormal basis of each eigenspace.

Perform Gram-Schmidt process to the vectors

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

we find

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

For E_3, we get

$$\vec{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Therefore, the orthogonal matrix

$$S = \begin{bmatrix} | & | & | \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$$

diagonalizes the matrix A:

$$S^{-1}AS = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

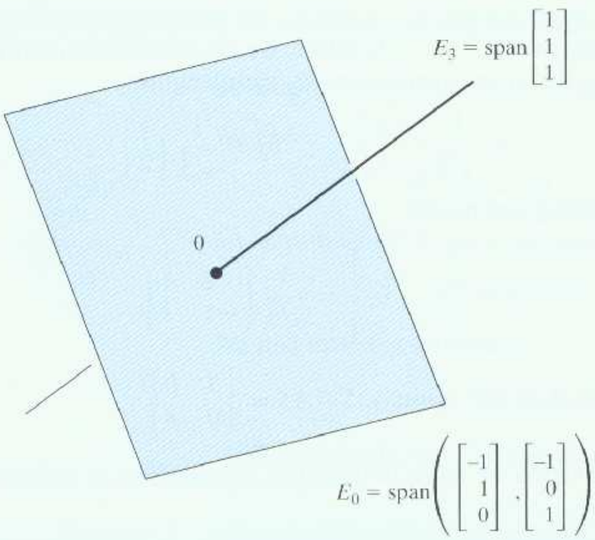


Figure 2 The eigenspaces E_0 and E_3 are orthogonal complements.

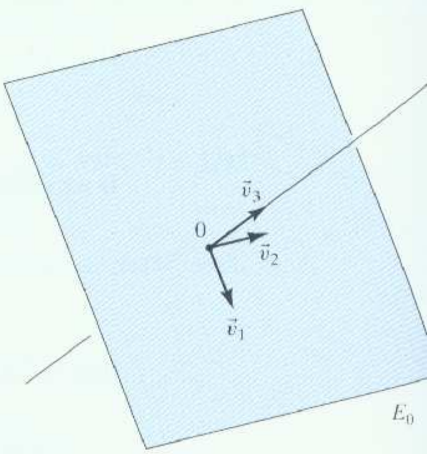


Figure 3

Algorithm 8.1.4 Orthogonal diagonalization of a symmetric matrix A

1. Find the eigenvalues of A, and find a basis of each eigenspace.

2. Using the Gram-Schmidt process, find an orthonormal basis of each eigenspace.

3. Form an orthonormal eigenbasis $\vec{u}_1, \vec{u}_2, ..., \vec{u}_n$ for A by combining the vectors you find in the last step, and let

$$P = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & ... & \vec{u}_n \\ | & | & & | \end{bmatrix}$$

P is orthogonal, and $P^{-1}AP$ will be diagonal.

Spectral Decomposition

Suppose that $A = PDP^{-1}$, where the columns of P are orthonormal eigenvectors $\vec{u}_1, \vec{u}_2, ..., \vec{u}_n$ of A and the corresponding eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$A = PDP^T = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 \vec{u}_1 & \cdots & \lambda_n \vec{u}_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} = \lambda_1 \vec{u}_1 \vec{u}_1^T + \cdots + \lambda_n \vec{u}_n \vec{u}_n^T$$

This representation of A is called a spectral decomposition of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A. Each term is an $n \times n$ matrix of rank 1. Furthermore, each matrix $\vec{u}_j \vec{u}_j^T$ is a projection matrix onto the subspace spanned by $\vec{u}_j$.

Example 4 Consider an invertible symmetric 2×2 matrix A. Show that the linear transformation $T(\vec{x} = A\vec{x}$ maps the unit circle into an ellipse, and find the lengths of the semimajor and the seminor axes of the ellipse in terms of the eigenvalues of A.

Solution

The spectral theorem tells us there is an orthonormal eigenbasis u_1, u_2 for T, with associated real eigenvalues λ_1, λ_2. Suppose that $|\lambda_1| > |\lambda_2|$. These eigenvalues will be nonzero, since A is invertible. The unit circle consists of all vectors of the form

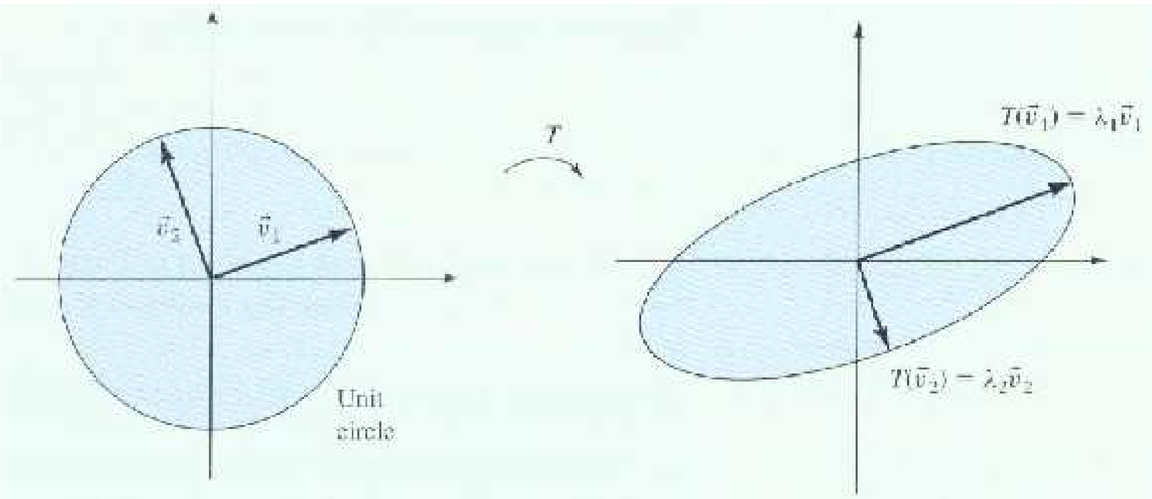
$$\vec{v} = \cos(t)u_1 + \sin(t)u_2$$

. The image of the unit circle will be

$$T(\vec{v}) = \cos(t)T(u_1) + \sin(t)T(u_2)$$

$$= \cos(t)\lambda_1 u_1 + \sin(t)\lambda_2 u_2$$

an ellipse whose semimajor axis has the length $\|\lambda_1 u_1\| = |\lambda_1|$, while the length of the semiminor axis is $\|\lambda_2 u_2\| = |\lambda_2|$. (See Figure 4).



8.2 Quadratic Forms

Example 1 Consider the function

$$q(x_1, x_2) = 8x_1^2 - 4x_1x_2 + 5x_2^2$$

Determine whether $q(0,0)$ is the global minimum.

Solution based on matrix technique
Rewrite

$$q\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 8x_1^2 - 4x_1x_2 + 5x_2^2$$

$$= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 8x_1 - 2x_2 \\ -2x_1 + 5x_2 \end{bmatrix}$$

Note that we split the contribution $-4x_1x_2$ equally among the two components.

More succinctly, we can write

$$q(\vec{x}) = \vec{x} \cdot A\vec{x}, \ \ where \ \ A = \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix}$$

or

$$q(\vec{x}) = \vec{x}^T A\vec{x}$$

The matrix A is symmetric by construction. By the spectral theorem, there is an orthonormal eigenbasis $\vec{v}_1, \vec{v}_2$ for A. We find

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

with associated eigenvalues $\lambda_1 = 9$ and $\lambda_2 = 4$.

Let $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2$, we can express the value of the function as follows:

$$q(\vec{x}) = \vec{x} \cdot A\vec{x} = (c_1\vec{v}_1 + c_2\vec{v}_2) \cdot (c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2)$$

$$= \lambda_1 c_1^2 + \lambda_2 c_2^2 = 9c_1^2 + 4c_2^2$$

Therefore, $q(\vec{x}) > 0$ for all nonzero $\vec{x}$. $q(0,0) = 0$ is the global minimum of the function.

Def 8.2.1 Quadratic forms
A function $q(x_1, x_2, \ldots, x_n)$ from R^n to R is called a quadratic form if it is a linear combination of functions of the form $x_i x_j$. A quadratic form can be written as

$$q(\vec{x}) = \vec{x} \cdot A\vec{x} = \vec{x}^T A\vec{x}$$

for a symmetric $n \times n$ matrix A.

Example 2 Consider the quadratic form

$q(x_1, x_2, x_3) = 9x_1^2 + 7x_2^2 + 3x_3^2 - 2x_1x_2 + 4x_1x_3 - 6x_2x_3$
Find a symmetric matrix A such that $q(\vec{x}) = \vec{x} \cdot A\vec{x}$ for all $\vec{x}$ in R^3.

Solution As in Example 1, we let
$a_{ii} =$ (coefficient of x_i^2),
$a_{ij} = \frac{1}{2}$ (coefficient of $x_i x_j$), if $i \neq j$.
Therefore,

$$A = \begin{bmatrix} 9 & -1 & 2 \\ -1 & 7 & -3 \\ 2 & -3 & 3 \end{bmatrix}$$

Change of Variables in a Quadratic Form

Fact 8.2.2 Consider a quadratic form $q(\vec{x}) = \vec{x} \cdot A\vec{x}$ from R^n to R. Let B be an orthonormal eigenbasis for A, with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$q(\vec{x}) = \lambda_1 c_1^2 + \lambda_2 c_2^2 + \ldots + \lambda_n c_n^2,$$

where the c_i are the coordinates of $\vec{x}$ with respect to B.

Let $x = Py$, or equivalently, $y = P^{-1}x = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, if change of variable is made in a quadratic form $x^T Ax$, then

$$x^T Ax = (Py)^T A(Py) = y^T P^T APy = y^T (P^T AP)y$$

Since P orghogonally diagonalizes A, the $P^T AP = P^{-1}AP = D$.

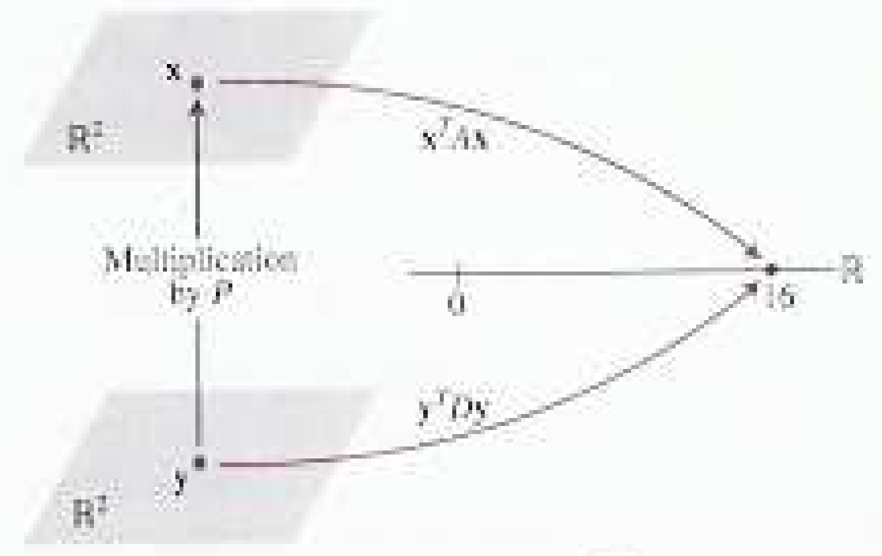
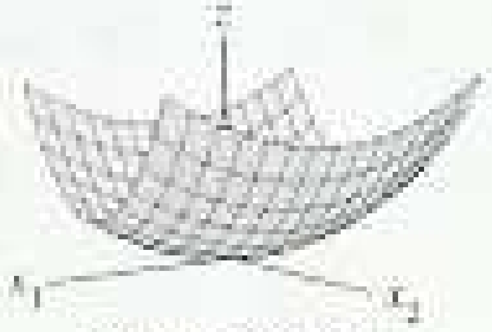


FIGURE 1 Change of variable in x^TAx.

Classifying Quadratic Form

Positive definite quadratic form
If $q(\vec{x}) > 0$ for all nonzero $\vec{x}$ in R^n, we say A is positive definite.
If $q(\vec{x}) \geq 0$ for all nonzero $\vec{x}$ in R^n, we say A is positive semidefinite.
If $q(\vec{x})$ takes positive as well as negative values, we say A is indefinite.



Positive Definite

Negative Definite

Indefinite

Example 3 Consider $m \times n$ matrix A. Show that the function $q(\vec{x}) = ||A\vec{x}||^2$ is a quadratic form, find its matrix and determine its definiteness.

Solution $q(\vec{x}) = (A\vec{x}) \cdot (A\vec{x}) = (A\vec{x})^T(A\vec{x}) = \vec{x}^T A^T A\vec{x} = \vec{x} \cdot (A^T A\vec{x})$.
This shows that q is a quadratic form, with symmetric matrix $A^T A$.
Since $q(\vec{x}) = ||A\vec{x}||^2 \geq 0$ for all vectors $\vec{x}$ in R^n, this quadratic form is positive semidefinite.
Note that $q(\vec{x}) = 0$ iff $\vec{x}$ is in the kernel of A. Therefore, the quadratic form is positive definite iff $ker(A) = \{\vec{0}\}$.

Fact 8.2.4 Eigenvalues and definiteness
A symmetric matrix A is positive definite iff all its eigenvalues are positive.

The matrix is positive semidefinite iff all of its eigenvalues are positive or zero.

Fact: The Principal Axes Theorem
Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $x = Py$, that transforms the quadratic form $x^T Ax$ into a quadratic form $y^T Dy$ with no cross-product term.

Principle Axes

When we study a function $f(x_1, x_2, \ldots, x_n)$ from R^n to R, we are often interested in the solution of the equation

$$f(x_1, x_2, \ldots, x_n) = k,$$

for a fixed k in R, called the level sets of f.

Example 4 Sketch the curve

$$8x_1^2 - 4x_1x_2 + 5x_2^2 = 1$$

Solution In Example 1, we found that we can write this equation as

$$9c_1^2 + 4c_2^2 = 1$$

where c_1 and c_2 are the coordinates of $\vec{x}$ with respect to the orthonormal eigenbasis

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

for $A = \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix}$. We sketch this ellipse in Figure 4.

The c_1-axe and c_2-axe are called the `principle axes` of the quadratic form $q(x_1, x_2) = 8x_1^2 - 4x_1x_2 + 5x_2^2$. Note that these are the eigenspaces of the matrix

$$A = \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix}$$

of the quadratic form.

Constrained Optimization

When a quadratic form Q has no cross-product terms, it is easy to find the maximum and minimum of $Q(\vec{x})$ for $\vec{x}^T\vec{x}x = 1$.

Example 1 Find the maximum and minimum values of $Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraint $\vec{x}^T\vec{x}x = 1$.

Solution

$$Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2 \le 9x_1^2 + 9x_2^2 + 9x_3^2$$

$$= 9(x_1^2 + x_2^2 + x_3^2) = 9$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. $Q(\vec{x}) = 9$ when $\vec{x} = (1, 0, 0)$. Similarly,

$$Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2 \ge 3x_1^2 + 3x_2^2 + 3x_3^2$$

$$= 3(x_1^2 + x_2^2 + x_3^2) = 3$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. $Q(\vec{x}) = 3$ when $\vec{x} = (0, 0, 1)$.

THEOREM Let A be a symmetric matrix, and define

$$m = min\{x^T Ax : \|\vec{x}\} = 1\}, M = max\{x^T Ax : \|\vec{x}\} = 1\}.$$

Then M is the greatest eigenvalues λ_1 of A and m is the least eigenvalue of A. The value of $x^T Ax$ is M when x is a unit eigenvector u_1 corresponding to eigenvalue M. The value of $x^T Ax$ is m when x is a unit eigenvector corresponding to m.

Proof

Orthogonally diagonalize A, i.e. $P^T AP = D$ (by change of variable $x = Py$), we can transform the quadratic form $x^T Ax = (Py)^T A(Py)$ into $y^T Dy$. The constraint $\|x\| = 1$ implies $\|y\| = 1$ since $\|x\|^2 = \|Py\|^2 = (Py)^T Py = y^T P^T Py = y^T(P^T P)y = y^T y = 1$.

Arrange the columns of P so that $P = \left[\begin{array}{ccc} u_1 & \cdots & u_n \end{array}\right]$ and $\lambda_1 \geq \cdots \geq \lambda_n$.

Given that any unit vector y with coordinates $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, observe that

$$y^T D y = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2$$

$$\geq \lambda_1 c_1^2 + \cdots + \lambda_1 c_n^2 = \lambda_1 \|y\| = \lambda_1$$

Thus $x^T A x$ has the largest value $M = \lambda_1$ when $y = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$, i.e. $x = Py = u_1$.

A similar argument show that m is the least eigenvalue λ_n when $y = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$, i.e. $x = Py = u_n$.

THEOREM Let A, λ_1 and u_1 be as in the last theorem. Then the maximum value of $x^T Ax$ subject to the constraints

$$x^T x = 1, x^T u_1 = 0$$

is the second greatest eigenvalue, λ_2, and this maximum is attained when x is an eigenvector u_2 corresponding to λ_2.

THEOREM Let A be a symmetric $n \times n$ matrix with an orthogonal diagonalization $A = PDP^{-1}$, where the entries on the diagonal of D are arranged so that $\lambda_1 \geq \cdots \geq \lambda_n$, and where the columns of P are corresponding unit eigenvectors $u_1, ..., u_n$. Then for $k = 2, ..., n$, the maximum value of $x^T Ax$ subject to the constraints

$$x^T x = 1, x^T u_1 = 0, ..., x^T u_{k-1} = 0$$

is the eigenvalue λ_k, and this maximum is attained when $x = u_k$.

The Singular Value Decomposition

The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks certain the eigenvectors. If $Ax = \lambda x$ and $x^T x = 1$, then

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

based on the diagonalization of $A = PDP^{-1}$.

The description has an analogue for rectangular matrices that will lead to the singular value decomposition $A = QDP^{-1}$.

Example If $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$, then the linear transformation $T(x) = Ax$ maps the unit sphere $\{x : \|x\| = 1\}$ in R^3 into an ellipse in R^2 (see Fig. 1). Find a unit vector at which $\|Ax\|$ is maximized.

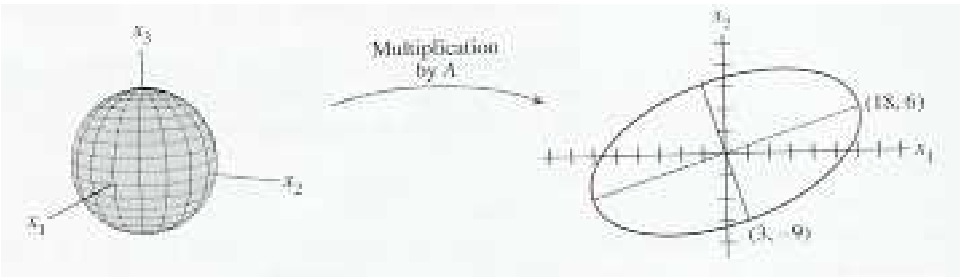


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

Observe that

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

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 to maximize the quadratic form $x^T(A^T A)x$ subject to the constraint $\|x\| = 1$.

Compute

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

Find the eigenvalues of $A^T A$: $\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0$, and the corresponding unit eigenvectors,

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

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

The Singular Values of an $m \times n$ Matrix

Let A be an $m \times n$ matrix. Then $A^T A$ is symmetric and can be orthogonally diagonalized. Let $\{v_1, ..., v_n\}$ be an orthonormal basis for R^n consisting of eigenvectors of $A^T A$, and let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues of $A^T A$. Then for $1 \leq i \leq n$,

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

So the eigenvalues of $A^T A$ are all nonnegative. Let

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$$

The singular values of A are the square roots of the eigenvalues of $A^T A$, denoted by $\sigma_1, ..., \sigma_n$. That is $\sigma_i = \sqrt{\lambda_i}$ for $1 \leq i \leq n$. The singular values of A are the lengths of the vectors $Av_1, ..., Av_n$.

Example

Let A be the matrix in the last example. Since the eigenvalues of A^TA are 360, 90, and 0, the singular values of A are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \sigma_2 = \sqrt{90} = 3\sqrt{10}, \sigma_3 = 0$$

Note that, the first singular value of A is the maximum of $\|Ax\|$ over all unit vectors, and the maximum is attained at the unit eigenvector v_1. The second singular value of A is the maximum of $\|Ax\|$ over all unit vectors that are orthogonal to v_1, and this maximum is attained at the second unit eigenvector, v_2. Compute

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

$$Av_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}$$

The fact that Av_1 and Av_2 are orthogonal is no accident, as the next theorem shows.

THEOREM Suppose that $\{v_1, ..., v_n\}$ is an orthonormal basis of R^n consisting of eigenvectors of $A^T A$, arranged so that the corresponding eigenvalues of $A^T A$ satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$, and suppose that A has r nonzero singular values. Then $\{Av_1, ..., Av_r\}$ is an orthogonal basis for im(A), and rank(A)=r.

Proof Because v_i and v_j are orthogonal for $i \neq j$,

$$(Av_i)^T(Av_j) = v_i^T A^T A v_j = v_i^T \lambda_j v_j = 0$$

Thus $\{Av_1, ..., Av_n\}$ is an orthogonal set. Furthermore, $Av_i = 0$ for $i > r$. For any y in im(A), i.e. $y = Ax$

$$y = Ax = A(c_1 v_1 + \cdots + c_n v_n)$$

$$= c_1 A v_1 + \cdots + c_r A v_r + 0 + \cdots + 0$$

Thus y is in Span$\{Av_1, ..., Av_r\}$, which shows that $\{Av_1, ..., Av_r\}$is an (orthogonal) basis for im(A). Hence rank(A)=dim im(A)=r.

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 $A^T A$, 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

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

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

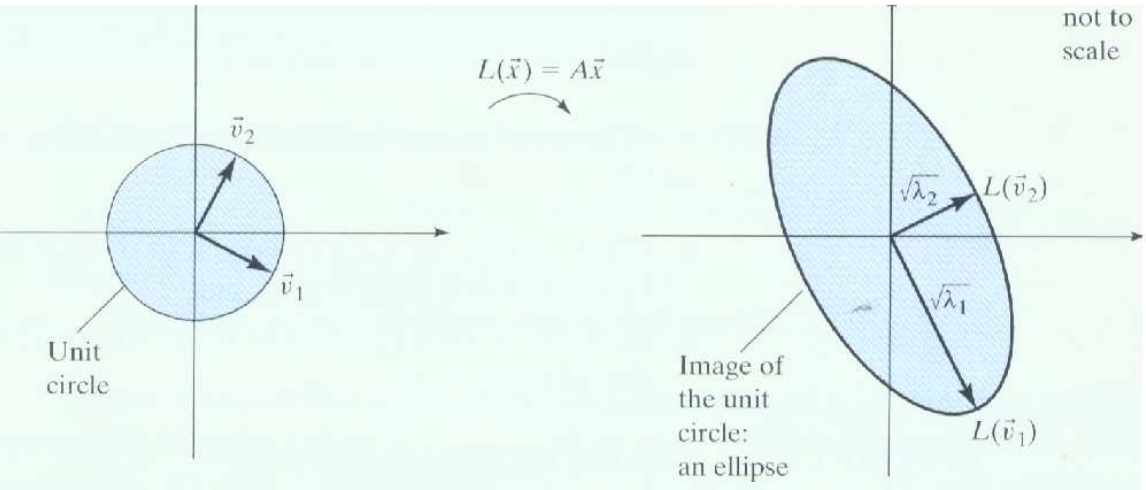
We can also compute $L(\vec{v}_1)$ and $L(\vec{v}_2)$ directly:

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

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

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.

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

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.

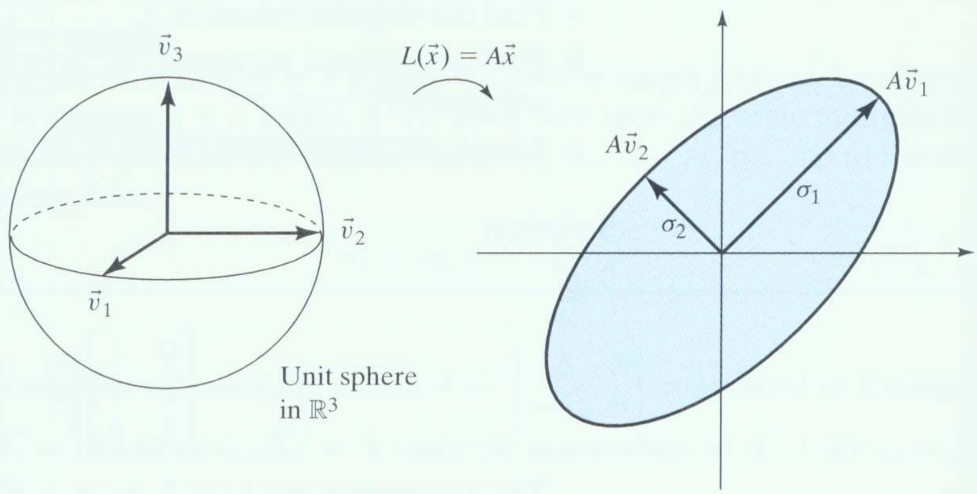

$\vec{v}_3$
$L(\vec{x}) = A\vec{x}$
$A\vec{v}_1$
$A\vec{v}_2$
σ_1
σ_2
$\vec{v}_2$
$\vec{v}_1$
Unit sphere
in $\mathbb{R}^3$

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:

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

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

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

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

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

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

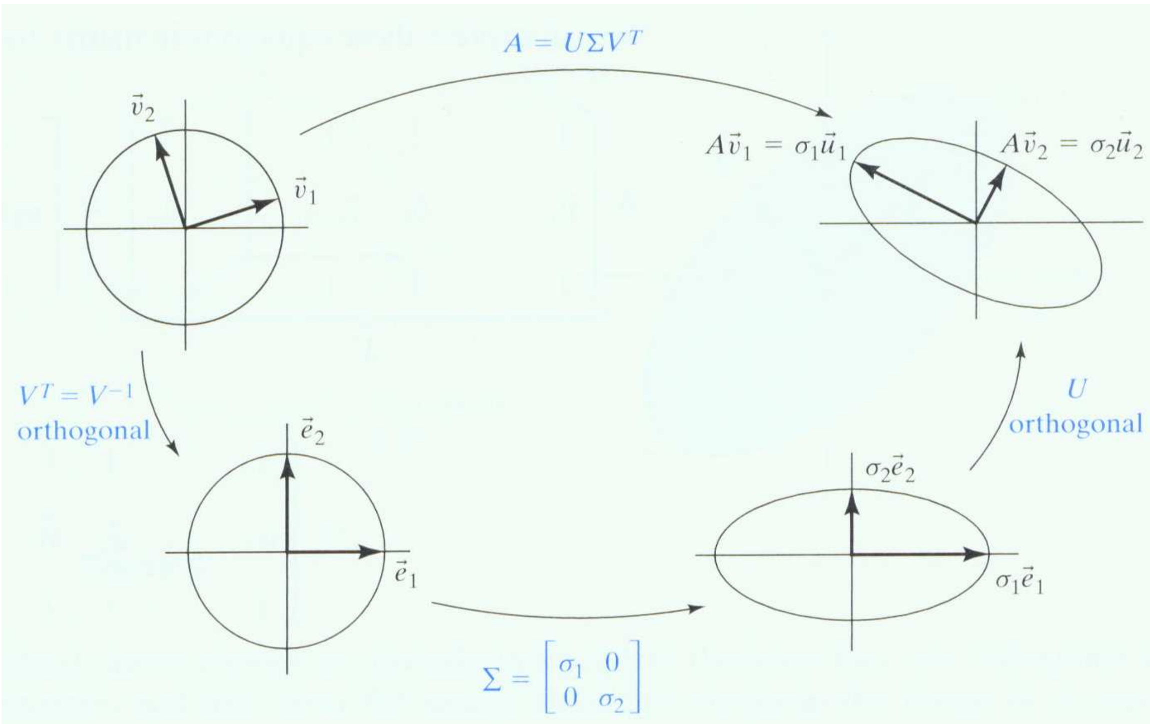
where $\vec{u}_i$ and $\vec{v}_i$ are the columns of U and V, respectively.

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

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

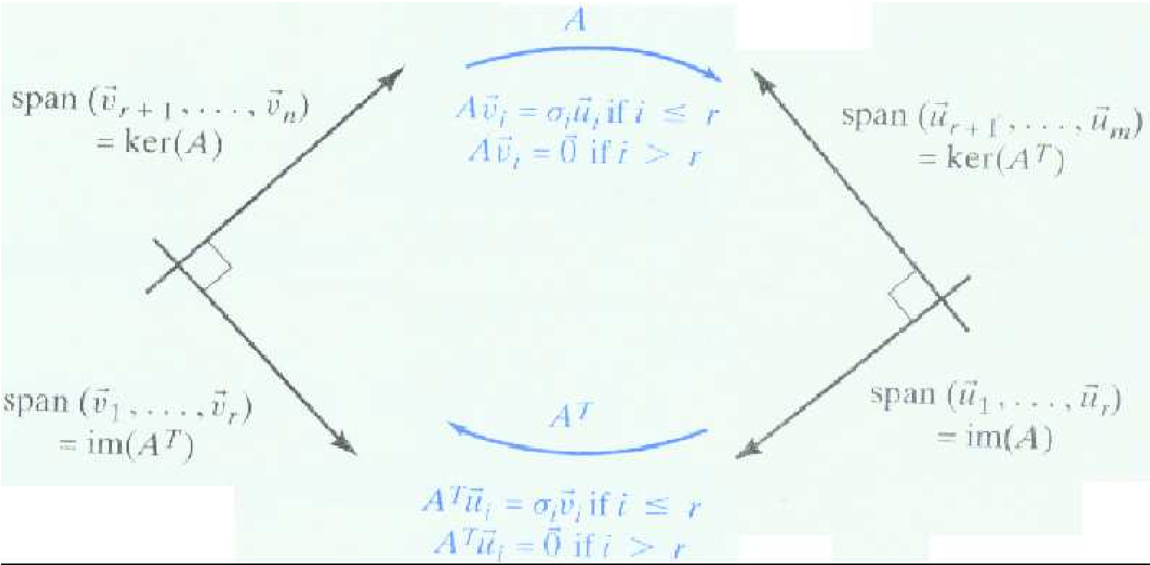
As before

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

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.

Application to Data Compression

Suppose a satellite transmits a picture containing 1000×1000 pixels. If the color of each pixel is digitized, this information can be represented in a 1000×1000 matrix A.

Suppose we know an SVD

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

Even if the rank r of the matrix A is large, most of the singular values will typically be very small (relatively to σ_1). If we neglect those, we get a good approximation $A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T + \ldots + \sigma_s \vec{u}_s \vec{v}_s^T$, where s is much smaller than r.

For example, if we choose $s = 10$, we need to transmit only the 20 vectors $\sigma_1 \vec{u}_1, \ldots, \sigma_{10} \vec{u}_{10}$ and $\vec{v}_1, \ldots, \vec{v}_{10}$ in R^{1000}, that is, 20,000 numbers.

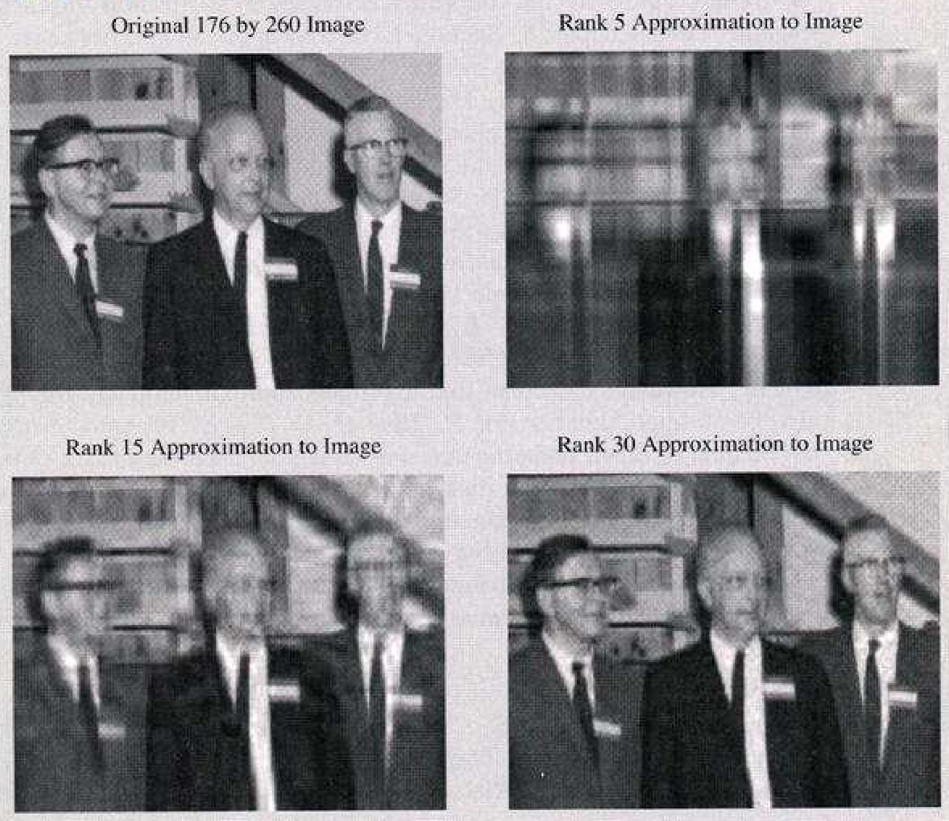


FIGURE 6.5.1 Courtesy Oakridge National Laboratory

Application to Information Retrieval

Consider the problem of searching a database for documents. If there are m possible key words and a total of n documents. Then the database can be represented by a $m \times n$ matrix A.

Two of the main problems are polysemy (words having multiple meanings) and synonymy (multiple words having the same meaning).

If we think of our database as an approximation. Some of the entries may contain extraneous components due to polysemy, and some may miss including components because of synonymy.

Suppose it were possible to correct for these problems and come up with a perfect database matrix P. Let $E = A - P$, then $A = P + E$.

We can think of E as a matrix representing the errors.

Latent semantic indexing (LSI)
The idea of LSI is that the lower-rank matrix may still provide a good approximation to P and, may actually involve less error.

The lower-rank approximation can be obtained by truncating the outer product expansion of the singular value decomposition of A. This is equivalent to setting

$$\sigma_{s+1} = \sigma_{s+2} = \ldots = \sigma_n = 0$$

and then setting $A_s = U_s \Sigma_s V_s^T$, the compact form of the singular value decomposition.

Speedup
The matrix vector multiplication $A^T \vec{q}$ requires a total of mn scalar multiplications.

On the other hand, $A_s^T = V_s \Sigma_s U_s^T$, and the multiplication $A_s^T \vec{q} = V_s(\Sigma_s(U_s^T \vec{q}))$ requires a total of $s(m+n+1)$ scalar multiplications.

Reference

S. J. Leon, Linear algebra with applications, 6th Ed., Prentice Hall. 2002.

Applications to Statistics

`Matrix of observations`

An example of two-dimensional data is given by a set of weights and heights of N college students. Let X_j denote the observation vector in R^2 that lists the weight and height of the jth student. Then, the matrix of observation has the form

$$\begin{array}{c} \begin{bmatrix} w_1 & w_2 & \dots & w_N \\ h_1 & h_2 & \dots & h_N \end{bmatrix} \\ \begin{array}{cccc} \uparrow & \uparrow & & \uparrow \\ X_1 & X_2 & \dots & X_N \end{array} \end{array}$$

`Mean and Covariance`

To prepare for principle component analysis, let $\begin{bmatrix} X_1 & \dots & X_N \end{bmatrix}$ be a $p \times N$ matrix of observations. The sample mean, M, of the observation vectors is given by

$$M = \frac{1}{N}(X_1 + \ldots + X_N)$$

Let

$$\hat{X}_k = X_k - M$$

The columns of the $p \times N$ matrix

$$B = \left[\begin{array}{cccc} \hat{X}_1 & \hat{X}_2 & \ldots & \hat{X}_N \end{array} \right]$$

have a zero sample mean, and B is said to be in `mean-deviation form`.

The `(sample) covariance matrix` is the $p \times N$ matrix S defined by

$$S = \frac{1}{N-1} BB^T$$

The entries s_{jj} is called the `variance` of x_j.
The `total variance` of the data is the sum of the variances on the diagonal of S, $totalvariance = trace(S)$.
The entries s_{ij} for $i \neq j$ is called the `covariance` of x_i and x_j.

Principle Component Analysis

Assume that the matrix $X = \begin{bmatrix} X_1 & \ldots & X_N \end{bmatrix}$ is already in mean-deviation form. The goal of principle component analysis is to find an orthogonal $p \times p$ matrix $P = \begin{bmatrix} u_1 & \ldots & u_p \end{bmatrix}$ that determines a change of variable, $X = PY$, or

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \ldots & u_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

such that the new variables $y_1, y_2, \ldots, y_p$ are uncorrelated and are arranged in order of decreasing variance.

Let $S = \frac{1}{N-1}XX^T$ be the covariance matrix of X. Since the covariance matrix of $Y = \begin{bmatrix} Y_1 & \ldots & Y_N \end{bmatrix}$ is $\frac{1}{N-1}YY^T = \frac{1}{N-1}(P^TX)(P^TX)^T = P^TSP$. So the desired orthogonal matrix P is one that makes P^TSP diagonal.

Let D be a diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_p$ of S on the diagonal, arranged that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$, and let P be an orthogonal matrix whose columns are the corresponding unit eigenvectors $u_1, \ldots, u_p$. Then $P^T SP = D$ and $S = PDP^T$.

The unit eigenvectors $u_1, \ldots, u_p$ are called the **`principle components`** of the data. The first principle component u_1 determines the new variable y_1 in the following way. Let $c_1, \ldots, c_p$ be the entries in u_1. Since u_1^T is the first row of P^T, the equation $Y = P^T X$ shows that

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_p^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

$$y_1 = u_1^T X = c_1 x_1 + c_2 x_2 + \ldots + c_p x_p$$

Thus, y_1 is a linear combination of the original variables $x_1, x_2, \ldots, x_p$, using the entries in the eigenvector u_1 as weights.

Reducing the Dimension

Principle component analysis is potentially valuable for applications in which most of the variation in the data is due to variations in only a few of the new variables, $y_1, y_2, \ldots, y_p$.

The variance of y_j is λ_j, and the quotient $\lambda_j / trace(S)$ measures the fraction of the total variance that is captured by y_j.

Reference

D. C. Lay, Linear algebra and its applications, 2nd Ed. Addison-Wesley, 2000.

Example The following table lists the weights and heights of five boys:

Boy	#1	#2	#3	#4	#5
$Weight$	120	125	125	135	145
$Height$	61	60	64	68	72

First, arrange the data in mean-deviation form. The sample mean vector is easily seen to be $M = \begin{pmatrix} 130 \\ 65 \end{pmatrix}$. Subtract M from the observation vectors and obtain

$$B = \begin{pmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{pmatrix}$$

Then the sample covariance matrix is

$$S = \frac{1}{5-1} \begin{pmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{pmatrix} \begin{pmatrix} -10 & -4 \\ -5 & -5 \\ -5 & -1 \\ 5 & 3 \\ 15 & 7 \end{pmatrix}$$

$$= \frac{1}{4} \begin{pmatrix} 400 & 190 \\ 190 & 100 \end{pmatrix} = \begin{pmatrix} 100 & 47.5 \\ 47.5 & 25 \end{pmatrix}$$

The eigenvalues of S are (to decimal places)

$$\lambda_1 = 123.02 \ and \ \lambda_2 = 1.98$$

The unit eigenvector corresponding to λ_1 is $u_1 = \begin{pmatrix} 0.900 \\ 0.436 \end{pmatrix}$. For the size index, set

$$y = 0.900\widehat{w} + 0.436\widehat{h}$$

where $\widehat{w}$ and $\widehat{h}$ are weight and height, respectively, in mean-deviation form. The variance of this index over the data set is 123.02. Because the total variance is tr(S)=100+25=125, the size index accounts for practically all (98.4%) of the variance of the data.