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

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

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

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 $P^T AP = P^{-1}AP = D$.

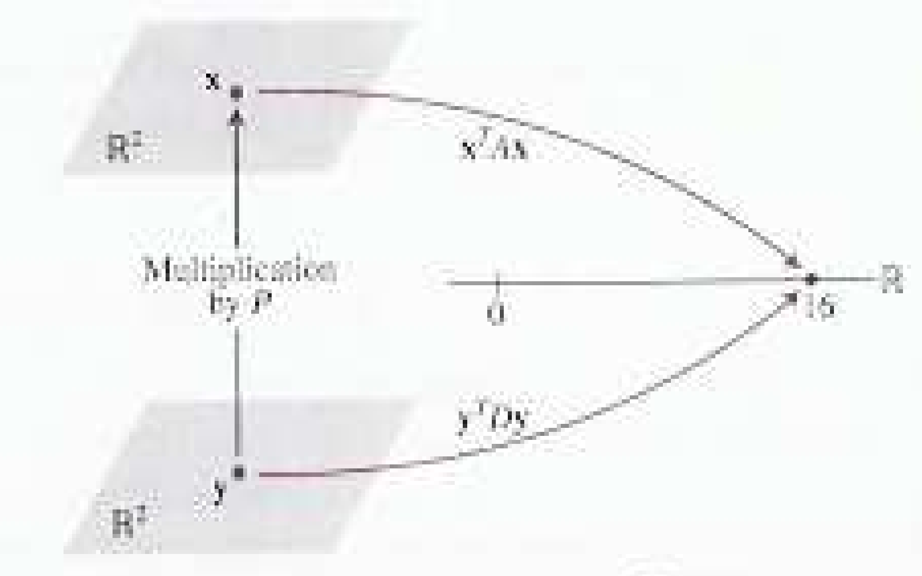


FIGURE 1 Change of variable in x^TAx.

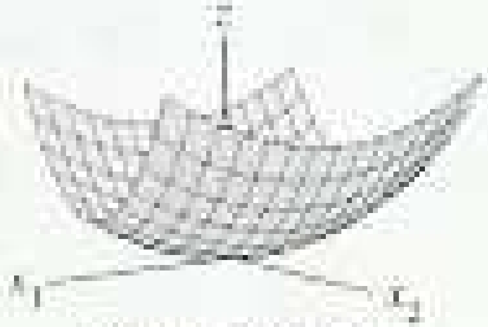
Classifying Quadratic Form

Positive definite quadratic form

If $q(\vec{x}) > 0$ for all nonzero $\vec{x}$ in R^n, we say A is positive definite.

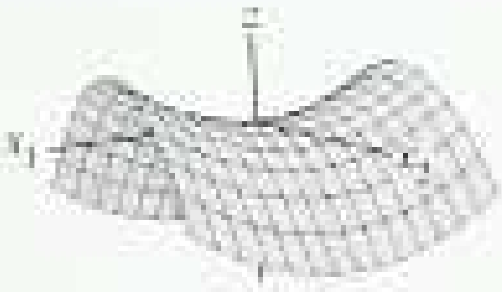
If $q(\vec{x}) \geq 0$ for all nonzero $\vec{x}$ in R^n, we say A is positive semidefinite.

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

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

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

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

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

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

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.

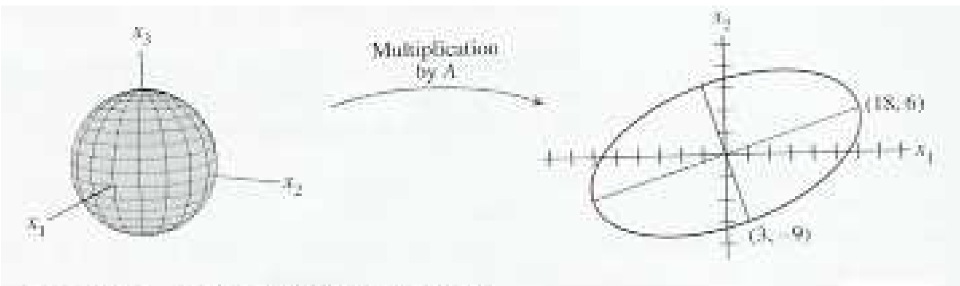


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

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

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

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 $1 \leq i \leq n$,

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

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.