1-1-1989

Introduction To Optimization Methods Part 1: Mathematical Review

Stanislaw H. Zak
Purdue University

Introduction To Optimization Methods

Part 1: Mathematical Review

Stanislaw H. Żak

TR-EE 89-1
January 1989

School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

INTRODUCTION TO
OPTIMIZATION METHODS

PART 1: MATHEMATICAL REVIEW

STANISLAW H. ŻAK
SCHOOL OF ELECTRICAL ENGINEERING
PURDUE UNIVERSITY
MSEE Bldg., Rm 233B
WEST LAFAYETTE
IN 47907
Ph #: (317) 494-6443

PART 1: MATHEMATICAL REVIEW

The following is a review of some basic definitions, notations and relations from linear algebra, geometry, and calculus that will be used frequently throughout this book.

1. LINEAR ALGEBRA

The purpose of this chapter is to review those aspects of linear algebra and calculus, that will be of importance in the chapters to follow.

VECTOR SPACES

A column n-vector is defined as an array of n numbers which may be real or complex, and is denoted as follows

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} .$$

The number a_i is referred to as the ith component of the vector **a**. Similarly we can define a row n-vector as

$$\mathbf{a}^T = [a_1, a_2, ..., a_n] ,$$

where the superscript ^{T} denotes the transposition operation.

Two vectors $\mathbf{a} = [a_1, a_2, ..., a_n]^T$ and $\mathbf{b} = [b_1, b_2, ..., b_n]^T$ are equal if $a_i = b_i$, $i = 1, 2, ..., n$.

The sum of the vectors **a** and **b** denoted as **a**+**b** is the vector

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, a_2 + b_2, ..., a_n + b_n]^T .$$

The operation of addition of vectors has the following properties:

(i) The operation is commutative;

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} ,$$

(ii) The operation is associative;

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}) ,$$

(iii) There is a **o** vector

$$\mathbf{o} = [0,0,...,0]^T .$$

Note that

$$\mathbf{a} + \mathbf{o} = \mathbf{o} + \mathbf{a} = \mathbf{a} .$$

The following vector

$$[a_1 - b_1 \ , \ a_2 - b_2,...,a_n - b_n]^T$$

is called the difference between **a** and **b** and is denoted as $\mathbf{a} - \mathbf{b}$.

The vector $\mathbf{o} - \mathbf{b}$ is denoted as $-\mathbf{b}$ and is called an inverse vector of **b**. Note that

$$\begin{aligned} &\mathbf{b} + (\mathbf{a} - \mathbf{b}) = \mathbf{a} , \\ &-(-\mathbf{b}) = \mathbf{b} , \\ &-(\mathbf{a} - \mathbf{b}) = \mathbf{b} - \mathbf{a} . \end{aligned}$$

The vector $\mathbf{a} - \mathbf{b}$ is the unique solution of the vector equation

$$\mathbf{a} + \mathbf{x} = \mathbf{b} .$$

Indeed, suppose $\mathbf{x} = [x_1,x_2,...,x_n]^T$ is a solution to $\mathbf{a} + \mathbf{x} = \mathbf{b}$. Then

$$a_1 + x_1 = b_1 \ , \ \ a_2 + x_2 = b_2,...,a_n + x_n = b_n \ ,$$

hence

$$x_i = b_i - a_i \quad (i = 1,...,n) ,$$

and thus

$$\mathbf{x} = \mathbf{b} - \mathbf{a} .$$

We define an operation of scalar multiplication as follows

$$\alpha\mathbf{a} = [\alpha a_1, \alpha a_2, ..., \alpha a_n]^T \; .$$

This operation has the following properties:

(i) distributive laws hold

$$\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b} \; ,$$
$$(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a} \; ,$$

(ii) the operation is associative

$$\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a} \; ,$$

(iii) there is an identity such that

$$1\mathbf{a} = \mathbf{a} \; ,$$

(iv) $\alpha\mathbf{o} = \mathbf{o}$,

(v) $0\mathbf{a} = \mathbf{o}$,

(vi) $(-1)\mathbf{a} = -\,\mathbf{a}$.

Note that

$$\alpha\mathbf{a} = \mathbf{o}$$

if and only if $\alpha = 0$ or $\mathbf{a} = \mathbf{o}$. Indeed $\alpha\mathbf{a} = \mathbf{o}$ is equivalent to

$$\alpha a_1 = \alpha a_2 = ... = \alpha a_n = 0 \; .$$

If $\alpha = 0$ or $\mathbf{a} = 0$ then $\alpha\mathbf{a} = \mathbf{o}$. If $\mathbf{a} \neq 0$ then at least one of its components $a_k \neq 0$. For this component $\alpha a_k = 0$, hence we have to have $\alpha = 0$ and similar arguments can be applied to the case when $\alpha \neq 0$.

A set of nonvanishing vectors $\mathbf{a}_1,...,\mathbf{a}_k$ is said to be *linearly independent* if the equality

$$\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + ... + \alpha_k\mathbf{a}_k = \mathbf{o}$$

implies that all coefficients α_i $(\mathbf{i} = 1,...,k)$ are equal to zero.

The set of the nonvanishing vectors $\mathbf{a}_1,...,\mathbf{a}_k$ is *linearly dependent* if at least one α_i does not vanish.

Note that a set composed of a single vector $\mathbf{a} = \mathbf{o}$ is linearly dependent, for if $\alpha \neq 0$ then $\alpha\mathbf{a} = \alpha\mathbf{o} = \mathbf{o}$.

A set composed of a single nonvanishing vector $\mathbf{a} \neq \mathbf{o}$ is linearly independent since

$$\alpha\mathbf{a} = \mathbf{o} \quad \text{implies} \quad \alpha = 0 \, .$$

A vector $\mathbf{a}$ is said to be a *linear combination* of vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k$ if there is a set of numbers $\alpha_1, ..., \alpha_k$ such that

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + ... + \alpha_k \mathbf{a}_k \, .$$

A set of vectors $\mathbf{a}_1, ..., \mathbf{a}_k$ is linearly dependent if and only if one of the vectors from the set is a linear combination of the remaining vectors.

Indeed, if $\mathbf{a}_1, ..., \mathbf{a}_k$ are linearly dependent then

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + ... + \alpha_k \mathbf{a}_k = \mathbf{o}$$

where at least one of the coefficients $\alpha_i \neq 0$.

If $\alpha_i \neq 0$ then

$$\mathbf{a}_i = -\frac{\alpha_1}{\alpha_i}\mathbf{a}_1 - \frac{\alpha_2}{\alpha_i}\mathbf{a}_2 - ... - \frac{\alpha_k}{\alpha_i}\mathbf{a}_k \, .$$

On the other hand if

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + ... + \alpha_k \mathbf{a}_k$$

then

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + ... + \alpha_k \mathbf{a}_k + (-1)\mathbf{a} = \mathbf{o}$$

where the last coefficient is nonzero, and thus the set of vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k, \mathbf{a}$ is linearly dependent.

The set of all column vectors (n-tuples) $\mathbf{a} = [a_1, a_2, ..., a_n]^T$ whose components a_i's are real numbers is called the *real vector space* and is denoted $\mathbb{R}^n$.

Let $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k$ be arbitrary vectors from $\mathbb{R}^n$. Then the set of all their linear combinations is referred to as a *linear subspace* spanned by these vectors and denoted as

$$\text{Span}\ [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k]\ .$$

Note that a subspace Span $[\mathbf{a}]$ is composed of the vectors $\alpha\mathbf{a}$, where α is an arbitrary real number ($\alpha\in\mathbb{R}$).

Also observe that if $\mathbf{a}$ is linearly dependent upon $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k$ then

$$\text{Span}\ [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k, \mathbf{a}] = \text{Span}\ [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k]\ .$$

Every linear subspace contains a zero vector, for if $\mathbf{a}$ is an element of the subspace so is $(-1)\mathbf{a} = -\mathbf{a}$. hence $\mathbf{a} - \mathbf{a} = \mathbf{o}$ also belongs to the subspace.

Note that a subspace V of $\mathbb{R}^n$ is a set that is closed under the operations of vector addition and scalar multiplication. For if $\mathbf{a}$ and $\mathbf{b}$ are vectors in V then the vector $\alpha\mathbf{a} + \beta\mathbf{b}$ is also in V for every pair of scalars α and β.

We have the following property of linear subspaces.

If V is a linear subspace and the vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k$ are linearly independent and $\mathbf{a}_i\in V$, $i = 1, ..., k$ where k is the maximal number of such vectors then

$$V = \text{Span}\ [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k]\ .$$

To prove this statement note that any vector $\mathbf{a}$ of V is linearly dependent on $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k$, hence

$$\mathbf{a} = \alpha_1\mathbf{a}_1 + ... + \alpha_k\mathbf{a}_k$$

which implies

$$V \subset \text{Span}\ [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_k]\ .$$

On the other hand the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ belong to V, therefore

$$V \supset \mathrm{Span}\ [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]\ .$$

Hence

$$V = \mathrm{Span}\ [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]\ .$$

Any set of linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ such that

$$V = \mathrm{Span}\ [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$$

is referred to as a *basis* of the subspace V, and the number k is the dimension of the subspace.

If the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ form a basis of V then any vector $\mathbf{a}$ of V can be represented in the unique way as a linear combination of the basis vectors

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_k \mathbf{a}_k\ .$$

where $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, k$.

To prove uniqueness of the representation of $\mathbf{a}$ in terms of the basis vectors assume that

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_k \mathbf{a}_k$$

and

$$\mathbf{a} = \beta_1 \mathbf{a}_1 + \beta_2 \mathbf{a}_2 + \ldots + \beta_k \mathbf{a}_k\ .$$

Hence

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a} + \ldots + \alpha_k \mathbf{a}_k = \beta_1 \mathbf{a}_1 + \beta_2 \mathbf{a}_2 + \ldots + \beta_k \mathbf{a}_k\ ,$$

or

$$(\alpha_1 - \beta_1)\mathbf{a}_1 + (\alpha_2 - \beta_2)\mathbf{a}_2 + \ldots + (\alpha_k - \beta_k)\mathbf{a}_k = \mathbf{o}\ .$$

Since $\mathbf{a}_i (i = 1, \ldots, k)$ are linearly independent we have to have

$$\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \ldots = \alpha_k - \beta_k = \mathbf{0} \ ,$$

that is

$$\alpha_i = \beta_i \qquad (i = 1,\ldots,k)$$

which proves the uniqueness of the representation.

If $k \geq n+1$ then the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ ($\mathbf{a}_i \in \mathbb{R}^n$) are linearly dependent, that is there exist scalars $\beta_1, \ldots, \beta_k$ such that

$$\sum_{i=1}^{k} \beta_i^2 > 0$$

and

$$\sum_{i=1}^{k} \beta_i \mathbf{a}_i = \mathbf{0} \ .$$

If $k \geq n+2$ then

$$\sum_{i=1}^{k} \beta_i^2 > 0$$

and

$$\sum_{i=1}^{k} \beta_i = \mathbf{0} \ .$$

To prove the above introduce the following vectors [3]

$$\bar{\mathbf{a}}_i = [1, a_1, a_2, \ldots, a_n]^T \in \mathbb{R}^{n+1} \ , \quad i = 1,\ldots,k, \qquad k \geq n+2 \ .$$

Since any $n+2$ vectors in $\mathbb{R}^{n+1}$ are linearly dependent, there exist scalars β_i $(i = 1,\ldots,k)$ such that

$$\sum_{i=1}^{k} \beta_i \ \bar{\mathbf{a}}_i = \mathbf{0}$$

and

$$\sum_{i=1}^{k} \beta_i^2 > 0 .$$

It follows from $\sum_{i=1}^{k} \beta_i \bar{\mathbf{a}}_i = \mathbf{0}$ and the fact that the first component of each $\bar{\mathbf{a}}_i$ is equal to one that

$$\sum_{i=1}^{k} \beta_i = 0 .$$

The *natural basis for* $\mathbb{R}^n$ is the set of vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \dots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} .$$

The reason for calling these vectors the natural basis is that

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 e_1 + x_2 e_2 + \dots + x_n e_n .$$

Sometimes it is convenient to change the natural basis $e_1, e_2, \dots, e_n$ to a new basis $f_1, f_2, \dots, f_n$. One then needs to the able to express the column x' of new coordinates in terms of the column x of old coordinates. We have

$$Fx' = [f_1, f_2, \dots, f_n] x' = x_1' f_1 + x_2' f_2 + \dots + x_n' f_n = x .$$

RANK OF A MATRIX

A *matrix* is a rectangular array of numbers. A matrix with m rows and n columns is called an m×n matrix. For example

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} .$$

Let us denote the kth column of A by $\mathbf{a}_k$. Hence

$$\mathbf{a}_k = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} .$$

The maximal number of linearly independent columns of A is called the rank of the matrix, and denoted rank A. We will show the following:

The rank of a matrix A is invariant under the following operations:

(i) multiplication of the columns of A by nonzero scalars,

(ii) interchange of the columns,

(iii) adding a linear combination of other columns to a given column.

Indeed, let $\mathbf{b}_k = \alpha_k \mathbf{a}_k$ where $\alpha_k \neq \mathbf{0}$, k = 1,...,n, Let B = $[\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n]$. Obviously

$$\text{Span}\,[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] = \text{Span}\,[\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n] \ ,$$

and thus

$$\text{rank } A = \text{rank } B \ .$$

In order to prove (ii), note that the number of linearly independent vectors does not depend on their order. To prove (iii) let

$$\begin{aligned} \mathbf{b}_1 &= \mathbf{a}_1 + c_2\mathbf{a}_2 + \dots + c_n\mathbf{a}_n \ , \\ \mathbf{b}_2 &= \mathbf{a}_2 \ , \\ &\vdots \\ \mathbf{b}_n &= \mathbf{a}_n \ , \end{aligned}$$

and let $B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n]$. Obviously

$$\text{Span } [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n] \subset \text{Span } [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \ .$$

On the other hand

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{b}_1 - c_2\mathbf{b}_2 - \dots - c_n\mathbf{b}_n \ , \\ \mathbf{a}_2 &= \mathbf{b}_2 \ , \\ &\vdots \\ \mathbf{a}_n &= \mathbf{b}_n \ . \end{aligned}$$

Hence

$$\text{Span } [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \subset \text{Span } [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n] \ .$$

Therefore

$$\text{rank } A = \text{rank } B \ .$$

With each square (m = n) matrix there is associated scalar called the *determinant* of the matrix. We shall denote the determinant of the square matrix A by det A or by $\Delta(A)$.

The determinant of a square matrix is a function of its columns and it has the following properties [8]:

(i) The determinant of the matrix $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ is a linear function of each column, that is

$$\Delta(\mathbf{a}_1, \dots, \mathbf{a}_{k-1}, \alpha\mathbf{a}_k' + \beta\mathbf{a}_k'', \mathbf{a}_{k+1}, \dots, \mathbf{a}_n)$$

$$= \alpha\Delta(\mathbf{a}_1, \dots, \mathbf{a}_{k-1}, \mathbf{a}_k', \mathbf{a}_{k+1}, \dots, \mathbf{a}_n)$$

$$+\beta\Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}''_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_n)\ ,$$

(ii) If for some k $\quad \mathbf{a}_k = \mathbf{a}_{k+1}$ $\quad$ then

$$\Delta(\mathbf{a}_1,\dots,\mathbf{a}_k,\mathbf{a}_{k+1},\dots\mathbf{a}_n) = \mathbf{0}\ ,$$

(iii) If

$$\mathbf{a}_1 = \begin{bmatrix}1\\0\\0\\\vdots\\0\end{bmatrix},\quad \mathbf{a}_2 = \begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix},\dots,\quad \mathbf{a}_n = \begin{bmatrix}0\\0\\0\\\vdots\\1\end{bmatrix}$$

then

$$\Delta(\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_n) = 1\ .$$

Note that if $\alpha = \beta = 0$ in (i) then

$$\Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{o},\mathbf{a}_{k+1},\dots,\mathbf{a}_n) = 0\ .$$

Thus if one of the column vanishes then the determinant is equal to zero.

Furthermore the determinant does not change its value if we add to a column another column multiplied by a scalar. The above follows from (i) and (ii) [8];

$$\Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k + \alpha\mathbf{a}_{k\mp\ 1},\mathbf{a}_{k+1},\dots,\mathbf{a}_n)$$

$$= \Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_n)$$

$$+\alpha\Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_{k\mp\ 1},\mathbf{a}_{k+1},\dots,\mathbf{a}_n)$$

$$= \Delta(\mathbf{a}_1,\dots,\mathbf{a}_n)\ .$$

However, the determinant changes its sign if we interchange neighboring columns. To show this property note the following [8]:

$$\Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_n)$$

$$= \Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k + \mathbf{a}_{k+1},\mathbf{a}_{k+1},\dots,\mathbf{a}_n)$$

$$= \Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k + \mathbf{a}_{k+1},\mathbf{a}_{k+1} - (\mathbf{a}_k + \mathbf{a}_{k+1}),\dots,\mathbf{a}_n)$$

$$= \Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k + \mathbf{a}_{k+1}, - \mathbf{a}_k,\dots,\mathbf{a}_n)$$

$$= \Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},(\mathbf{a}_k + \mathbf{a}_{k+1}) - \mathbf{a}_k,-\mathbf{a}_k,\dots,\mathbf{a}_n)$$

$$\overset{(i)}{=} - \Delta(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_{k+1},\mathbf{a}_k,\dots,\mathbf{a}_n) .$$

The (m-i)th order *minor* M of a m×n ($m \leq n$) matrix A is the determinant of the matrix obtained from A by deleting i rows and i + (n−m) columns.

One can use minors to investigate the rank of a matrix. In particular, we have:

If in an m×n ($m \geq n$) matrix A there exists a minor of the nth order, then the columns of A are linearly independent, that is rank A = n.

On the contrary, suppose that the columns of A are linearly dependent. Then there are scalars $x_i (i = 1,\dots,n)$ such that

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n = \mathbf{0} ,$$

and $\sum\limits_{i=1}^{n} x_i^2 > 0$.

The above vector equality is equivalent to the following set of m equations

$$\begin{array}{llll} a_{11}x_1 & + a_{12}x_2 & + \dots & + a_{1n}x_n = 0 \\ a_{21}x_1 & + a_{22}x_2 & + \dots & + a_{2n}x_n = 0 \\ \vdots & \vdots & & \vdots \\ a_{n1}x_1 & + a_{n2}x_2 & + \dots & + a_{nn}x_n = 0 \\ \vdots & \vdots & & \vdots \\ a_{m1}x_1 & + a_{m2}x_2 & + \dots & + a_{mn}x_n = 0 . \end{array}$$

Assume that the minor

$$M = \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \det \left[\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \ldots, \bar{\mathbf{a}}_n\right] \neq 0 \ .$$

From the properties of determinants it follows that the columns $\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \ldots, \bar{\mathbf{a}}_n$ are linearly independent. Hence the columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ have to be linearly independent too.

From the above it follows that if there is a nonzero minor then the column associated with this nonzero minor are linearly independent.

If a matrix A has an rth order minor M with the properties

(i) $M \neq 0$,

(ii) any minor of A which is formed by adding a row and a column to M is zero, then

$$\text{rank } A = r \ .$$

Thus the rank of a matrix is equal to its highest order nonzero minor.

A *nonsingular* matrix is a square matrix whose determinant is non-zero.

Suppose that A is a matrix. Then there is another matrix B such that

$$AB = BA = I$$

if and only if A is nonsingular. We call this matrix B the *inverse matrix* to A and write $B = A^{-1}$.

LINEAR EQUATIONS

Suppose we are given m equations in n unknowns of the form

$$\begin{array}{lllll} a_{11}x_1 & + a_{12}x_2 & + \dots & + a_{1n}x_n & = b_1 \\ a_{21}x_1 & + a_{22}x_2 & + \dots & + a_{2n}x_n & = b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1}x_1 & + a_{m2}x_2 & + \dots & + a_{mn}x_n & = b_m . \end{array} \tag{1.1}$$

We can represent the above set of equations as follows

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n = \mathbf{b} , \tag{1.2}$$

where

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Associated with the system of equations (1.1) are the following matrices

$$A = \left[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\right],$$

and

$$[A \,|\mathbf{b}] = \left[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \;|\mathbf{b}\right].$$

Note that (1.1) can also be represented as

$$A\mathbf{x} = \mathbf{b} , \tag{1.3}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Theorem 1.1.

The system of equations (1.1) has a solution if and only if

$$\text{rank } A = \text{rank } [A \,|\mathbf{b}] \,. \tag{1.4}$$

Proof:

Necessity ($\Rightarrow$):

Suppose the system (1.1) has a solution. Therefore **b** is a linear combination of the columns of A, that is (1.2) holds. From the above it follows that **b** belongs to the space spanned by the columns of A.

Sufficiency ($\Leftarrow$):

Suppose that (1.4) holds. Let rank A = rank $[A \,|\mathbf{b}] = r$. Thus we have r linearly independent columns of A. Let $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_r$ be these columns. The matrix $[A \,|\mathbf{b}]$ has also r linearly independent columns and they are $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_r$. The remaining columns of $[A \,|\mathbf{b}]$ can be expressed as linear combinations of $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_r$. In particular **b** can be expressed as a linear combination of these columns. Hence (1.2) holds.

□

Consider the equation $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{R}^{m \times n}$, and rank $A = m$.

Theorem 1.2.

All the solutions of $A\mathbf{x} = \mathbf{b}$, where rank $A = m$, can be obtained by assigning arbitrary values for $n-m$ variables and solving for the remaining ones.

Proof [8]:

We have rank A = m, therefore we can find m linearly independent columns of **A**. Let $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m$ be such columns. Rewrite equation (1.2) as follows

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + ... + x_m \mathbf{a}_m = \mathbf{b} - x_{m+1} \mathbf{a}_{m+1} + ... + x_n \mathbf{a}_n \ . \tag{1.5}$$

Assign to $x_{m+1}, x_{m+2}, ..., x_n$ arbitrary values, say

$$x_{m+1} = d_{m+1} \ , \ x_{m+2} = d_{m+2}, ..., x_n = d_n \ , \tag{1.6}$$

and let

$$B = [\mathbf{a}_1, \mathbf{a}_2, ... \mathbf{a}_m] \in \mathbb{R}^{m \times m} \ . \tag{1.7}$$

Note that det B $\neq$ 0.

We can represent (1.5) as follows

$$B \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \Big[\mathbf{b} - d_{m+1} \mathbf{a}_{m+1} + ... + d_n \mathbf{a}_n \Big] . \tag{1.8}$$

The matrix B is invertible, therefore we can solve for $[x_1, x_2, ..., x_m]^T$. Using (1.8) we can find all the solutions of $A\mathbf{x} = \mathbf{b}$. Indeed, if $\mathbf{x} = [d_1, ..., d_m, d_{m+1}, ..., d_n]^T$ is a solution to $A\mathbf{x} = \mathbf{b}$, then substituting (1.6) into (1.5) yields (1.8). But (1.8) has a unique solution for $x_1, ..., x_m$. Hence we have to have $x_1 = d_1, x_2 = d_2, ..., x_m = d_m$.

□

Corollary 1.2.

The system $A\mathbf{x} = \mathbf{b}$, $A \in \mathbb{R}^{m \times n}$, has a unique solution if and only if rank A = rank $[A \,|\mathbf{b}] = n$.

THE ABSOLUTE VALUE OF A NUMBER

The absolute value of a number a denoted $|a|$, is defined as follows:

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

The following formulae hold [6]:

(i) $|a| = |-a|$,

(ii) $-|a| \leq a \leq |a|$,

(iii) $|a+b| \leq |a| + |b|$,

(iv) $|a| - |b| \leq |a-b| \leq |a| + |b|$,

(v) $|ab| = |a| \cdot |b|$

(vi) $|a| \leq c$ and $|b| \leq d$ imply $|a+b| \leq c+d$,

(vii) the inequality $|a| < b$ is equivalent to $-b < a < b$, or to $a < b$ and $-a < b$.

THE PRINCIPLE OF INDUCTION

The principle may be stated as follows [6].

Assume that a given property of positive integers satisfies the following conditions:

(i) the number 1 possesses this property,

(ii) if the number n possesses this property then the number n+1 possesses it too.

The principle of induction states that under these assumption, any positive integer possesses the property.

The principle of induction is in agreement with the following intuitive argument. If the number 1 possesses the given property then, the second condition implies that the number 2 possesses the property. But then again the second condition implies that the number 3 possesses this property, etc. Thus the principle of induction gives a mathematical formulation of our intuitive reasoning.

UPPER BOUNDS

Consider a set S of real numbers. A number M is called an *upper bound* of S if

$$x \leq M \qquad \forall \ x \in S .$$

A set of real numbers that has an upper bound is said to be *bounded above*.

The Least Upper Bound Axiom

Every nonempty set S of real numbers that has an upper bound has a *least upper bound* or *supremum* and is denoted

$$\sup \{x \mid x \in S\} .$$

Examples

(i) $\sup\{\frac{1}{2} , \frac{2}{3} , \dots , \frac{n}{n+1} , \dots\} = 1$,

(ii) $\sup\{-\frac{1}{2} , -\frac{1}{8} , -\frac{1}{27} , \dots , -\frac{1}{n^3} , \dots\} = 0$,

(iii) $\sup\{x \mid x^2 < 3\} = \sup\{x \mid -\sqrt{3} < x < \sqrt{3}\} = \sqrt{3}$.

Theorem 1.3.

If $M = \sup\{x \mid x \in S\}$ and $\epsilon > 0$, then there is at least one number x in S such that

$$M - \epsilon < x \leq M .$$

Proof [9]:

The condition $x \leq M$ is satisfied by all numbers x in S by virtue of the Least Upper Bound Axiom. We have to show that for an $\epsilon > 0$ there is some number $x \in S$ such that

$$M - \epsilon < x .$$

Suppose on the contrary that there is no such number in S. Then

$$x \leq M - \epsilon \qquad \forall \; x \in S$$

and $M - \epsilon$ would be an upper bound of S that is less than M which is the least upper bound.

□

To illustrate the above Theorem consider the following set of real numbers

$$S = \left\{ \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots, \frac{n}{n+1}, \dots \right\}.$$

Note that

$$\sup\{x \mid x \in S\} = 1 .$$

Let $\epsilon = 0.0001$, then

$$1 - 0.0001 < x \leq 1$$

where for example

$$x = \frac{99999}{100000} .$$

LOWER BOUNDS

A number m is called a *lower bound* for S if

$$m \leq x \quad \forall \; x \in S \; .$$

Sets that have lower bounds are said to be *bounded below.*

Theorem 1.4.

Every nonempty set of real numbers that has a lower bound has a *greatest lower bound* or *infimum* denoted as

$$\inf \{x \mid x \in S\} \; .$$

Proof [9]:

By assumption S is nonempty and has a lower bound s. Thus

$$s \leq x \qquad \text{for all } x \in S \; .$$

Hence

$$-x \leq -s \qquad \text{for all } x \in S \; ,$$

that is

$$\{-x \mid x \in S\}$$

has an upper bound -s. From the Least Upper Bound Axiom we conclude that $\{-x : x \in S\}$ has a least upper bound (supremum) we call it s_0. Since

$$-x \leq s_0 \qquad \text{for all } x \in S \; ,$$

we have

$$-s_0 \leq x \qquad \text{for all } x \in S \; ,$$

and thus $-s_0$ is a lower bound for S. We now claim that $-s_0$ is the greatest lower bound (infimum) of the set S. Indeed, if there existed a number x_1 satisfying

$$-s_0 < x_1 \leq x \qquad \forall \; x \in S$$

then we would have

$$-x \leq -x_1 < s_0 \qquad \forall \; x \in S \, ,$$

and thus s_0 would not be the supremum of $\{-x \mid x \in S\}$.

□

Theorem 1.5.

If $m = \inf \{x \mid x \in S\}$ and $\epsilon > 0$, then there is at least one number x in S such that

$$m \leq x < m + \epsilon \, .$$

Proof

The proof is similar to the one of Theorem 1.3.

□

THE INTERMEDIATE-VALUE THEOREM

Lemma 1.1.

Let f be continuous on [a,b]. If $f(a) < 0 < f(b)$ or $f(b) < 0 < f(a)$, then there is a number c between a and b such that $f(c) = 0$.

Proof [9]:

Suppose that $f(a) < 0 < f(b)$. The other case can be treated in a similar fashion. We have $f(a) < 0$, thus from the continuity of f we know that there exists a number ξ such that f is negative on $[a,\xi)$. Let

$$c = \sup\{\xi \mid f < 0 \text{ on } [a,\xi)\} .$$

Clearly, $c \leq b$. Furthermore, we cannot have $f(c) > 0$, for then f would be positive on some interval extending to the left of c. From the properties of supremum (see Theorem 1.3), f is negative to the left of c. The above arguments imply that $c < b$. We cannot have $f(c) < 0$, for then there would be an interval $[a,t)$ with $t > c$, on which f is negative, and this would contradict the definition of c. Therefore $f(c) = 0$.

□

Theorem 1.6. (The Intermediate-Value Theorem)

If f is continuous on $[a,b]$ and C is a number between f(a) and f(b), then there is at least one number c between a and b for which $f(c) = C$.

Proof [9]:

Suppose

$$f(a) < C < f(b) .$$

The function

$$g(x) = f(x) - C$$

is continuous on $[a,b]$. We have

$$g(a) = f(a) - C < 0 \quad \text{and} \quad g(b) = f(b) - C > 0 .$$

From Lemma 1.1 there is a number c between a and b such that $g(c) = 0$. Hence

$f(c) = C$.

□

THE MAXIMUM-MINIMUM THEOREM
(THEOREM OF WEIERSTRASS)

Lemma 1.2.

If f is continuous on $[a,b]$, then f is bounded on $[a,b]$.

Proof [9]:

Consider the following set $\{x \mid x \in [a,b] \text{ and f is bounded on } [a,x]\}$. This set is nonempty and bounded above by b. Let

$$c = \sup \{x \mid \text{f is bounded on } [a,x]\} .$$

Now we shall show that $c = b$. Suppose that $c < b$. From the continuity of f at c it follows that f is bounded on $[c - \delta , c + \delta]$ for some $\delta > 0$. Being bounded on $[a , c - \delta]$ and on $[c - \delta , c + \delta]$, it is bounded on $[a , c + \delta]$. This contradicts our choice of c. We can therefore conclude that $c = b$. This means that f is bounded on $[a,x]$ for all $x < b$. From the continuity of f, we know that it is bounded on some interval $[b - \delta , b]$. Since $b - \delta < b$, f is bounded on $[a , b - \delta]$. Being bounded on $[a , b - \delta]$ and $[b - \delta , b]$, f is bounded on $[a,b]$.

□

Theorem 1.7. (The Maximum-Minimum Theorem)

If f is continuous on [a,b], then f takes on both a maximum value M and a minimum value m on [a,b].

Proof [9]:

By Lemma 1.2, f is bounded on [a,b]. Let

$$M = \sup \{f(x) \mid x \in [a,b]\} .$$

We will now show that there exists c in [a,b] such that $f(c) = M$. To accomplish this, we set

$$g(x) \triangleq \frac{1}{M - f(x)} .$$

If f does not take on the value of M, then g is continuous on [a,b] and thus by virtue of Lemma 1.2, bounded on [a,b]. But it is clear that g cannot be bounded on [a,b]. Thus the assumption that f does not take on the value M has led us to a contradiction. The other case that f takes on a minimum value m can be proved in a similar manner.

□

SEQUENCES

A sequence of numbers $a_1, a_2, ..., a_n, ...$ is a set of points $(1,a_1),(2,a_2),...,(n,a_n),....$

A sequence is *increasing* if $a_1 < a_2 < a_3 < ....$ In general a sequence is increasing if $a_n < a_{n+1}$. If $a_n \leq a_{n+1}$, then we say that the sequence is *nondecreasing.* Similarly, one can define *decreasing* and *nonincreasing* sequences. Increasing and/or decreasing sequences are called *monotone* sequences.

A number g is called the *limit* of the infinite sequence $a_1, a_2, ..., a_n, ...$ if for any positive ϵ there exists a number k such that for all $n > k$

$$|a_n - g| < \epsilon ,$$

that is a_n lies between $g-\epsilon$ and $g+\epsilon$ for all $n > k$. In other words; for any $\epsilon > 0$, $|a_n - g| < \epsilon$ is satisfied for all sufficiently large n's. (see Fig.)

We denote the limit of a sequence by

$$g = \lim_{n \to \infty} a_n .$$

A sequence which has a limit is called a *convergent* sequence. A sequence which has no limit is called *divergent.*

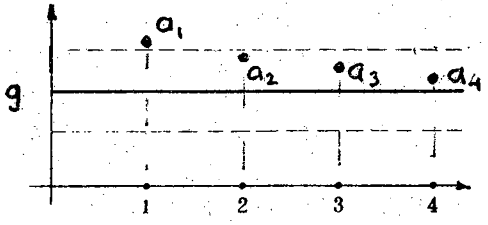


Fig. 1.1. Illustration of the notion f the limit.

We shall now show that if a limit exists then it is unique. We will prove this assertion by contradiction [6]. Let us assume that a sequence $a_1,a_2,...$ has two different limits, say g_1 and g_2. Then we have $|g_1 - g_2| > 0$. Let $\epsilon = \frac{1}{2} |g_1 - g_2|$. From the definition of a limit there exist k_1 and k_2 such that for $n > k_1$, $|a_n - g_1| < \epsilon$, and for $n > k_2$, $|a_n - g_2| < \epsilon$. Let $m = \max\{k_1,k_2\}$. Then if $n > m$, $|a_n - g_1| < \epsilon$ and $|a_n - g_2| < \epsilon$. If we now add $|a_n - g_1| < \epsilon$ and $|a_n - g_2| < \epsilon$ then we obtain

$$|a_n - g_1| + |a_n - g_2| < 2\epsilon .$$

In general

$$|a - b| \le |a| + |b| .$$

Hence

$$|-g_1 + g_2| \le |a_n - g_1| + |a_n - g_2| .$$

Therefore

$$|-g_1 + g_2| = |g_1 - g_2| < 2\epsilon .$$

But we assumed

$$|g_1 - g_2| = 2\epsilon .$$

Thus we have a contradiction and the proof is complete.

A sequence is bounded if there exists a number M such that $|a_n| < M$ for all n.

Theorem 1.8.

Every convergent sequence is bounded.

Proof [6]:

Let

$$g = \lim_{n \to \infty} a_n .$$

Choose $\epsilon = 1$. Then a number k exists such that

$$|a_n - g| < 1 \quad \text{for all } n > k .$$

In general

$$|a| - |b| \leq |a - b| ,$$

hence

$$|a_n| - |g| \leq |a_n - g| < 1 ,$$

and thus

$$|a_n| < |g| + 1 \quad \text{for all } n > k .$$

Let

$$M > \max\{|a_1|, \ldots, |a_k|, |g| + 1\} .$$

Obviously

$$M > |a_{k+i}| \ , \ i = 1,2,...$$

Therefore

$$M > |a_n| \quad \text{for all } n \ .$$

and thus the sequence is bounded.

□

Let a sequence $a_1, a_2, ..., a_n, ...$ and an increasing sequence of positive integers $m_1, m_2, ..., m_n, ...$ be given.

The sequence

$$b_1 = a_{m_1} \ , \ b_2 = a_{m_2}, ..., b_n = a_{m_n} \ , ...$$

is called a *subsequence* of the sequence $a_1, a_2, ..., a_n, ...$

In general, we have

$$m_n \geq n \ .$$

Indeed, this is obvious for $n=1$, that is $m_1 \geq 1$, since m_1 is a positive integer. We now apply the principle of induction. We assume that $m_n \geq n$ for a given n. Then we have $m_{n+1} > m_n \geq n$, hence $m_{n+1} \geq n+1$. Thus $m_n \geq n$ for all n.

One can say that every subsequence is obtained from the given sequence by neglecting a number of elements in this sequence. Hence a subsequence $\{a_{m_{k_n}}\}$ of a subsequence $\{a_{m_n}\}$ is a subsequence of the sequence $\{a_n\}$.

Theorem 1.9.

A subsequence of a convergent sequence is convergent to the same limit, that is if

$$\lim_{n\to\infty} a_n = g$$

and if

$$m_1 < m_2 < m_3 < \dots$$

then

$$\lim_{n\to\infty} a_{m_n} = g .$$

Proof [6]:

Let an $\epsilon > 0$ be given. Then a number $k > 0$ exists such that $|a_n - g| < \epsilon$ for any $n > k$. From $m_n \geq n$ it follows, $m_n > n > k$ and thus $|a_{m_n} - g| < \epsilon$ for any $m_n > n > k$. This means

$$\lim_{n\to\infty} a_{m_n} = g .$$

□

Theorem 1.10. (Bolzano-Weierstrass)

Every bounded sequence contains a convergent subsequence.

Proof [6]:

Let a sequence $a_1, a_2, \dots, a_n, \dots$ denoted as $\{a_n\}$ be bounded. Let **Z** be the set of all numbers x such that $x < a_n$, $i = 1,2,\dots$. The set **Z** is non-empty. Indeed, the number -M belongs to Z, for $|a_n| < M$, that is the inequality $-M < a_n$ holds for all n. The set Z is also bounded from above, for if $x \in Z$, then $x < M$ since if $x \geq M$ then the inequality $x < a_n$ would not be satisfied by any a_n.

Since the set **Z** is non-empty and bounded from above, the upper bound of this set exists. Denote this bound by g, that is

$$g = \sup \{x \,|\, x < a_n\} .$$

From Theorem 1.3 from the previous Section it follows that for any $\epsilon > 0$ there exist infinitely many a_n such that

$$g - \epsilon < a_n \leq g + \epsilon .$$

We will now show that g is the limit of a certain subsequence of the sequence $\{a_n\}$. This means that we have to define a sequence of positive integers $m_1 < m_2 < ...$ such that

$$\lim_{n \to \infty} a_{m_n} = g .$$

Let $\epsilon = 1$ in $g - \epsilon < a \leq g + \epsilon$. Thus

$$g - 1 < a_n \leq g + 1 .$$

Choose one of n's for which the above relation holds and denote it by m_1. We have

$$g - 1 < a_{m_1} \leq g + 1 .$$

Now let $\epsilon = \frac{1}{2}$. There exist infinitely many a_n such that

$$g - \frac{1}{2} < a_n \leq g + \frac{1}{2} .$$

Choose one of the elements for which the above relation holds and denote it by a_{m_2} where $m_1 < m_2$.

In general if m_n is defined, we choose m_{n+1} in such a way that

$$g - \frac{1}{n+1} < a_{m_{n+1}} \leq g + \frac{1}{n+1} \quad , \quad m_n < m_{n+1} .$$

Since $\lim\limits_{n \to \infty} \frac{1}{n} = 0$, we have

$$\lim_{n \to \infty} \left(g - \frac{1}{n}\right) = g = \lim_{n \to \infty} \left(g + \frac{1}{n}\right).$$

The two above relations imply

$$\lim_{n\to\infty} a_{m_n} = g \ .$$

The inequality $m_n < m_{n+1}$ on the other hand implies that the sequence $\{a_{m_n}\}$ is a subsequence of the sequence $\{a_n\}$.

□

Theorem 1.11. (Cauchy Theorem)

A sequence $\{a_n\}$ is convergent if and only if for every $\epsilon > 0$ there exist an r such that

$$|a_n - a_r| < \epsilon \quad \text{holds for all } n > r \ .$$

Proof [6]:

$\Rightarrow$ (if)

Let

$$\lim_{n\to\infty} a_n = g$$

and let $\epsilon > 0$ be given. Then for some r

$$|a_n - g| < \frac{1}{2}\,\epsilon \quad \text{for all } n \geq \ r \ .$$

In particular $|a_r - g| < \frac{1}{2}\,\epsilon$. Adding the above inequalities yields

$$|a_n - g| + |a_r - g| < \epsilon \ \text{ for all } \ n \geq \ r \ .$$

But

$$|a_n - a_r| < |a_n - g| + |a_r - g| \ .$$

Hence

$$|a_n - a_r| < \epsilon \quad \text{for all } n > r .$$

$\Leftarrow$ (only if)

Assume now that for every $\epsilon > 0$ an r exists such that

$$|a_n - a_r| < \epsilon \quad \text{for all } n > r .$$

We have to show that the above condition called the Cauchy condition implies

$$\lim_{n\to\infty} a_n = g .$$

First we show that $\{a_n\}$ is a bounded sequence. Let $\epsilon = 1$. Then an r exists such that $|a_n - a_r| < 1$ for all $n > r$. But

$$|a_n| - |a_r| \leq |a_n - a_r| < 1 ,$$

which means that

$$|a_n| < |a_r| + 1 .$$

Let

$$M > \max\{|a_1|,\ldots, |a_{r-1}|, |a_r| + 1\} .$$

Thus we have

$$|a_n| < M \quad \text{for all } n .$$

which implies that the sequence $\{a_n\}$ is bounded. By the Bolzano-Weierstrass theorem the sequence $\{a_n\}$ contains a convergent subsequence. Let

$$\lim_{n\to\infty} a_{m_n} = g , \quad m_1 < m_2 < \ldots .$$

We shall show that

$$\lim_{n\to\infty} a_n = g .$$

Let an $\epsilon > 0$ be given. The Cauchy condition implies that for some r

$$|a_n - a_r| < \frac{1}{3}\epsilon \quad \text{for all } n > r\,.$$

On the other hand $\lim\limits_{n\to\infty} a_{m_n} = g$ implies that for some k

$$|a_{m_n} - g| < \frac{1}{3}\epsilon \quad \text{for all } n > k\,.$$

One can select k so that $k > r$. If so then $|a_n - a_r| < \frac{1}{3}\epsilon$ and $|a_{m_n} - g| < \frac{1}{3}\epsilon$ for all $n > k$.

Since $m_n \geq n > r$

$$|a_{m_n} - a_r| < \frac{1}{3}\epsilon \text{ or } |a_r - a_{m_n}| < \frac{1}{3}\epsilon\,.$$

Adding $|a_n - a_r| < \frac{1}{3}\epsilon$, $|a_{m_n} - g| < \frac{1}{3}\epsilon$, and $|a_r - a_{m_n}| < \frac{1}{3}\epsilon$ yields

$$|a_n - a_r| + |a_{m_n} - a_r| + |a_{m_n} - g| < \epsilon\,.$$

But

$$|(a_n - a_r) + (a_r - a_{m_n}) + (a_{m_n} - g)| \leq |a_{n-r} - a_r| + |a_{m_n} - a_r| + |a_{m_n} - g| < \epsilon\,.$$

Thus

$$|a_n - g| < \epsilon \quad \text{for all } n > k$$

which implies

$$\lim_{n\to\infty} a_n = g\,.$$

□

2. ELEMENTARY CONCEPTS FROM GEOMETRY

SEGMENTS AND RAYS

All the analyses in the subsequent sections will be carried out in an n-dimensional space $\mathbb{R}^n$. The elements of this space are the n-component vectors $x = [x_1, x_2, \ldots, x_n]^T$. The vector $\overrightarrow{AB}$ with origin at $A(x_1, x_2, \ldots, x_n)$ and endpoint at $B(y_1, y_2, \ldots, y_n)$ has the components

$$[y_1 - x_1, y_2 - x_2, \ldots, y_n - x_n]^T .$$

Vectors in the space $\mathbb{R}^n$ can be added or multiplied by a scalar, by performing the corresponding operations on the components. For any three points A, B, and C in $\mathbb{R}^n$ we have

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC} .$$

It is easy to see that for any two points A and B

$$\overrightarrow{AB} = -\overrightarrow{BA} .$$

We now introduce the concept of a segment [2]. Let A and B be two distinct points in $\mathbb{R}^n$. Then we say that the point C lies on the *segment* AB if

$$\overrightarrow{AC} = \lambda \overrightarrow{AB} \tag{2.1}$$

where λ is a real number from the interval [0,1].

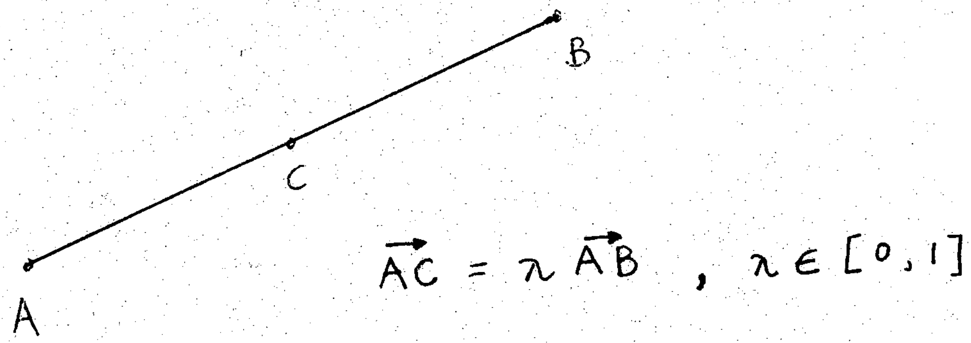


Fig. 2.1. Illustration of the concept of a segment.

Now, in addition to the points A, B, and C, we take an arbitrary point Q of $\mathbb{R}^n$. Without losing generality we stipulate Q = [0,0,...,0]. Then we have

$$\vec{AC} = \vec{QC} - \vec{QA}$$

and

$$\vec{AB} = \vec{QB} - \vec{QA} .$$

Hence the equation $\vec{AC} = \lambda\, \vec{AB}$ will take the form

$$\vec{QC} - \vec{QA} = \lambda\, (\vec{QB} - \vec{QA}) ,$$

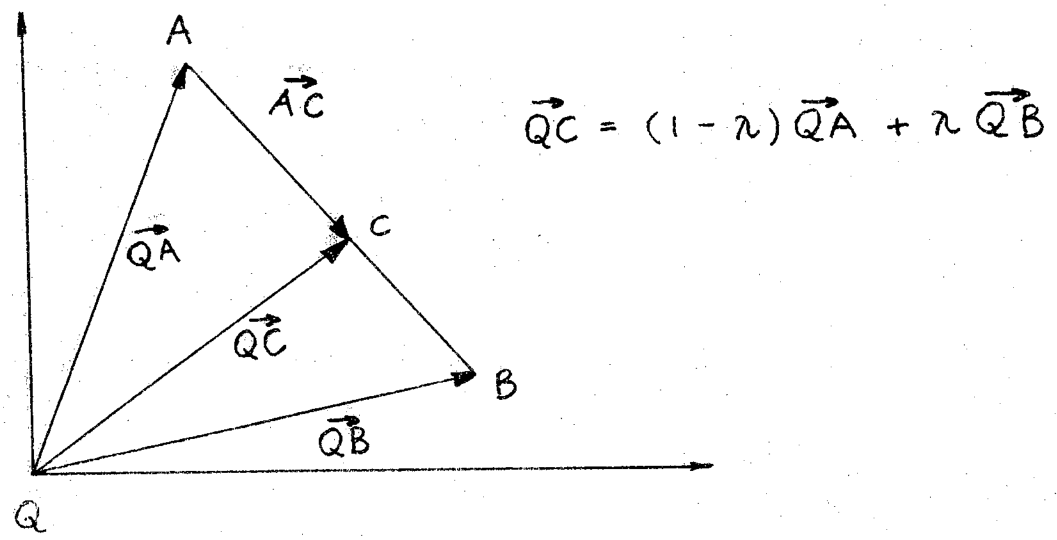


Fig. 2.2. An alternative illustration of the concept of a segment.

or

$$\overrightarrow{QC} = (1-\lambda)\,\overrightarrow{QA} + \lambda\,\overrightarrow{QB}\,. \tag{2.2}$$

In summary, the point C lies on the segment AB if and only if (2.1) or (2.2) holds where $\lambda \in [0,1]$ is a real number and point Q is arbitrary.

Let Q and A be two distinct points in $\mathbb{R}^n$. Then the set of all points B for which

$$\overrightarrow{QB} = \lambda\,\overrightarrow{QA} \quad , \quad \lambda \geq \; \mathbf{0}\,.$$

will be called the *ray* emanating from Q and passing through A [2].

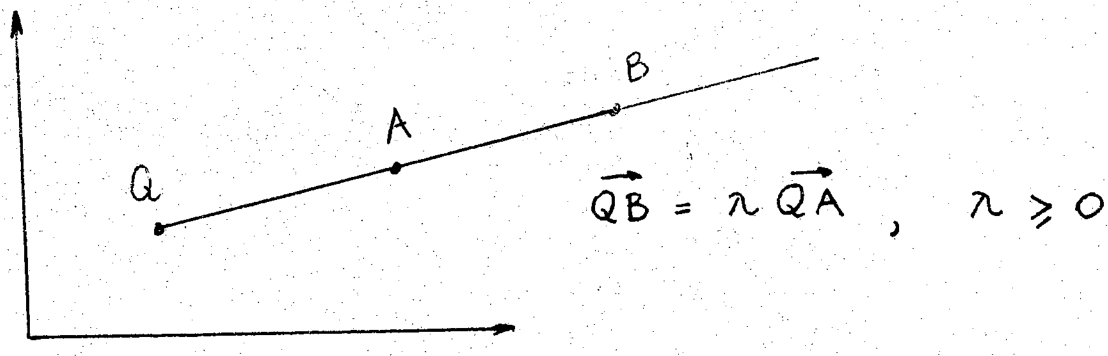


Fig. 2.3. Notion of the ray.

INNER PRODUCTS

For elements of *real* vector space we define an *inner product* $<x,y>$ to be any real valued function having the following properties:

Positivity: $<x,x> > 0$, except that $<0,0> = 0$,

Symmetry: $<x,y> = <y,x>$,

Additivity: $<x+y,z> = <x,z> + <y,z>$,

Homogeneity: $<rx,y> = r<x,y>$.

Although we assume additivity only in the first vector, we always have the properties of additivity and homogeneity in the second vector,

$$<x, y+z> = <x,y> + <x,z>$$

The above equation follows from the symmetry property of an inner product. Indeed

$$<x, y+z> = <y+z, x> = <y,z> + <z,x>$$

$$= <x,y> + <x,z>.$$

Similarly,

$$<x,ry> = <ry,x> = r<y,x> = r<x,y>.$$

The vectors x and y are said to be orthogonal if $<x,y> = 0$. The magnitude or norm of a vector x is

$$\|x\| = \sqrt{<x,x>}.$$

An n-dimensional space $\mathbb{R}^n$ equipped with the inner product will be called the Euclidean space and denoted $\mathbb{E}^n$.

Cauchy-Schwarz Inequality

For any two vectors x and y in $\mathbb{E}^n$, the Cauchy-Schwarz Inequality is true:

$$|<x,y>| \leq \|x\| \, \|y\|.$$

Proof [10]:

First assume that x and y are unit vectors, that is, that $\|x\| = \|y\| = 1$. Then,

$$\begin{aligned} 0 \leq \|x-y\|^2 &= <x-y,\ x-y> \\ &= \|x\|^2 - 2<x,y> + \|y\|^2 \\ &= 2 - 2<x,y>, \end{aligned}$$

or

$$<x,y> \leq 1.$$

Next, assuming that neither x nor y is zero (for the inequality obviously holds if one of them is zero), we can replace x and y by the unit vectors $x/\|x\|$ and $y/\|y\|$. The result is

$$<x,y> \leq \|x\| \, \|y\|.$$

Now replace x by -x to get

$$-<x,y> \leq \|x\| \, \|y\|.$$

The last two inequalities imply the absolute value inequality.

□

The magnitude or norm of a vector $\|x\|$ has the following properties:

Positivity: $\|x\| > 0$, except that $\|0\|=0$.

Homogeneity: $\|r\| = |r| \, \|x\|$, r real.

Triangle Inequality: $\|x+y\| \leq \|x\| + \|y\|$.

The following is a simple proof of the triangle inequality:

$$\|x+y\|^2 = \|x\|^2 + 2 < x,y > + \|y\|^2$$

(by Cauchy-Schwarz's Inequality)

$$\leq \|x\|^2 + 2 \, \|x\| \, \|y\| + \|y\|^2$$

$$+ (\|x\| + \|y\|)^2,$$

and therefore

$$\|x+y\| \leq \|x\| + \|y\|.$$

□

For elements of a *complex* vector space we define an *inner product* $< x,y >$ to be any complex valued function having the following properties:

1) $< x,x > > 0$, except that $< 0,0 > = 0$.

2) $< x,y > = \overline{< y,x >}$; the bar over the vectors denotes the complex conjugate.

3) $< x+y,z > = < x,z > + < y,z >$

4) $< rx,y > = r < x,y >$.

We can verify that in an n-dimensional complex space, the following definition of an inner product

$$< x,y > = \sum_{i=1}^{n} x_i \, \overline{y}_i$$

satisfies all four properties, where x_i and y_i are the ith components of x and y. One can deduce other properties, such as

$$< x, r_1 y + r_2 z > = \overline{r}_1 < x,y > + \overline{r}_2 < x,z >.$$

PROJECTORS

We know that a subspace V of $\mathbf{E}^n$ is a subset that is closed under the operations of vector addition and scalar multiplication. Thus if x_1 and x_2 are vectors in V then $\lambda x_1 + \mu x_2$ is also in V for every pair of scalars λ, μ. In other words, V is a subspace of $\mathbf{E}^n$ if $x_1, x_2 \in V \Rightarrow \quad \lambda x_1 + \mu x_2 \in V \ \forall \ \lambda , \mu \in \mathbb{R}$. Furthermore, the dimension of a subspace V is equal to the maximum number of linarly independent vectors in V. If V is a subspace of $\mathbf{E}^n$, then the orthogonal complement of V, denoted $V^\perp$, consists of all vectors that are orthogonal to every vector in V. Thus,

$$V^\perp = \{x \mid v^T x = 0 \quad \forall \ v \in V\} .$$

The orthogonal complement of V is a subspace, and together V and $V^\perp$ span $\mathbf{E}^n$ in the sense that every vector $x \in \mathbf{E}^n$ can be represented uniquely as

$$x = x_1 + x_2$$

where $x_1 \in V$, $x_2 \in V^\perp$. One can say that x_1 and x_2 are the orthogonal projections of x onto the subspaces V and $V^\perp$, respectively. In this case

$$\mathbf{E}^n = V \oplus V^\perp ,$$

i.e. $\mathbf{E}^n$ is a direct sum of V and $V^\perp$. Suppose that we are given a direct sum decomposition of $\mathbf{E}^n$ of the form

$$\mathbf{E}^n = V_1 \oplus V_2 .$$

Then every $x \in \mathbf{E}^n$ can be written uniquely as

$$x = x_1 + x_2 \ ; \ x_1 \in V_1 , \ x_2 \in V_2 .$$

Consider a mapping P of the form

$$Px \triangleq x_1 .$$

This defined mapping is a linear one. Indeed, let x , $x' \in \mathbf{E}^n$ such that

$$x = x_1 + x_2 \ , \ x' = x_1' + x_2' \ ,$$

where

$$x_1, x_1' \in V_1 \ , \ x_2, x_2' \in V_2 \ .$$

Then we have

$$Px = x_1 \quad \text{and} \quad Px' = x_1' \ .$$

Furthermore, for any scalars α , $\alpha' \in \mathbb{R}$ the following equality holds

$$\alpha x + \alpha' x' = (\alpha x_1 + \alpha' x_1') + (\alpha x_2 + \alpha' x_2')$$

where

$$\alpha x_1 + \alpha' x_1' \in V_1 \ , \quad \alpha x_2 + \alpha' x_2' \in V_2 \ .$$

Therefore

$$P(\alpha x + \alpha' x') = \alpha x_1 + \alpha' x_1' = \alpha Px + \alpha' Px' \ ,$$

and thus P is a linear mapping.

Note that if $x \in V_1$, then $x = x + \mathbf{0}$ where $x \in V_1$, $0 \in V_2$ and

$$Px = x \quad \forall \ x \in V_1 \ ,$$

which means that

$$PE^n = V_1 \ .$$

P is referred to as an orthogonal projector of $\mathbf{E}^n$ onto V_1 along V_2.

The projector P possesses the following property

$$P^2 = P \ .$$

Indeed, let $x \in E^n$ then $Px \in V_1$. Furthermore, $Px = x \ \forall \ x \in V_1$. Henceforth $P(Px) = Px$ for all $x \in E^n$.

Definition 2.1.

Transformations possessing the property $P^2 = P$ are called idempotent ones.

In the subsequent discussion we are going to utilize the following notation. Let $A \in \mathbb{R}^{m \times n}$. Let the range (or image) of A be denoted by

$$R(A) \triangleq \{y \in \mathbb{R}^m \mid y = Ax \quad \text{for} \quad \text{some } x \in \mathbb{R}^n\}$$

and let the null space (or kernel) of A be denoted by

$$N(A) \triangleq \{x \in \mathbb{R}^n \mid Ax = 0\} .$$

Theorem 2.1.

Every linear idempotent transformation is an orthogonal projector.

Proof

Let $P^2 = P$ and $V_1 = R(P)$ where V_1 is a linear subspace. Denote by V_2 a set of all the vectors of the form

$$x - Px$$

where

$$x \in E^n.$$

If $x', x'' \in V_2$ then $x' = x_0' - Px_0'$, $x'' = x_0'' - Px_0''$ where x_0', $x_0'' \in E^n$. So, for any α', $\alpha'' \in \mathbb{R}$ we have

$$\alpha' x' + \alpha'' x'' = \alpha' x_0' + \alpha'' x_0'' - (\alpha' P x_0' + \alpha'' P x_0'')$$

$$= (\alpha' x_0' + \alpha'' x_0'') - P(\alpha' x_0' + \alpha'' x_0'')$$

and thus $\alpha' x' + \alpha'' x'' \in V_2$. Thus indicates that V_2 is a linear subspace.

Furthermore, if $x_1 \in V_1$ then $Px_1 = x_1$, and

$$Px_1 = P(Px) = P^2x = Px = x_1 \ .$$

On the other hand, if $x_2 \in V_2$ then $Px_2 = 0$ since by definition of V_2 the vector $x_2 = x - Px$ where $x \in E^n$. Hence

$$Px_2 = Px - P(Px) = Px - P^2x = Px - Px = 0 \ .$$

i.e. $x_2 \in N(P)$. The conclusion is that any $x \in E^n$ can be represented as

$$x = x_1 + x_2 \ ; \ x_1 \in V_1 \ , \ x_2 \in V_2 \ .$$

This follows from the fact that

$$x = Px + (x - Px)$$

where $Px \in V_1$ and $(x - Px) \in V_2$. To complete the proof it must be shown that the above representation of x is unique. Note that if $x = x_1 + x_2$ then

$$Px = Px_1 + Px_2 = Px_1 + 0 = x_1$$

and thus

$$x_2 = x - x_1 = x - Px = (I - P)x \ ,$$

i.e. $x_2 \in R(I - P)$. Therefore, we see that x_1 and x_2 are uniquely determined by x. The above statement means that

$$E^n = V_1 \oplus V_2$$

and the proof is complete.

□

HYPERPLANES

The set of all points $x = [x_1, x_2, ..., x_n]^T$ which satisfy the linear equation

$$\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n = \beta \tag{2.3}$$

in which at least one of the coefficients α_i $(i = 1,...,n)$ differs from zero, is called a *hyperplane* of the space $\mathbb{E}^n$. The hyperplane may not be regarded as a subspace of $\mathbb{E}^n$ since, in general, it does not contain the origin. For n=2, the equation of the hyperplane has the form

$$\alpha_1 x_1 + \alpha_2 x_2 = \beta ,$$

which is the equation of a straight line. Thus straight lines are hyperplanes in $\mathbb{E}^2$. In $\mathbb{E}^3$ (three-dimensional space) the hyperplanes are ordinary planes. Thus one may speak of the dimension of the hyperplane. Note, that by translating a hyperplane so that it contains the origin of $\mathbb{E}^n$, it becomes a subspace of $\mathbb{E}^n$. The dimension of this subspace is n-1. Thus a hyperplane in $\mathbb{E}^n$ has dimension n-1.

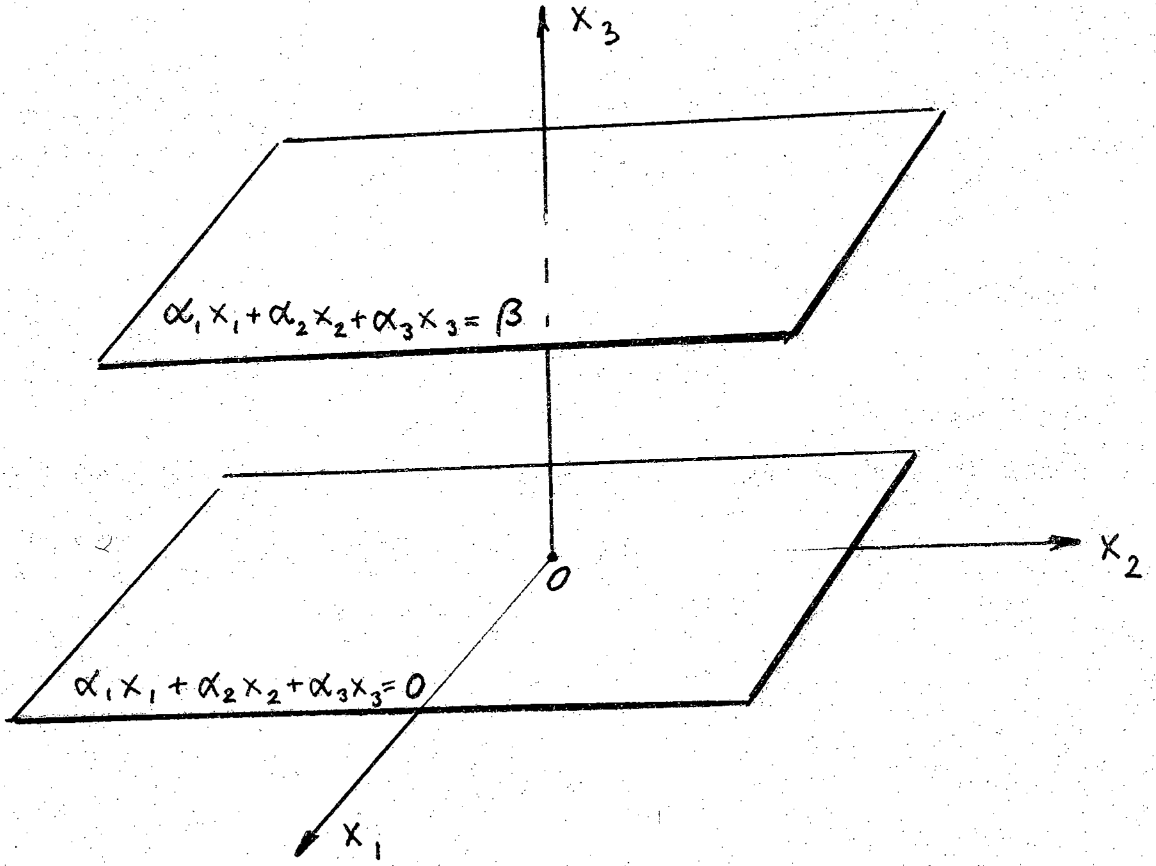

x_3
$\alpha_1 x_1+\alpha_2 x_2+\alpha_3 x_3=\beta$
x_2
0
$\alpha_1 x_1+\alpha_2 x_2+\alpha_3 x_3=0$
x_1

Fig. 2.4. Translation of a hyperplane.

The hyperplane (2.3) divides $\mathbb{E}^n$ into two half-spaces. One of these half-spaces consists of the points satisfying the inequality

$$\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n \geq \beta$$

denoted by

$$H_+ = \{x \in \mathbb{E}^n \mid \alpha^T x \geq \beta\},$$

where

$$\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]^T \ ,$$
$$x = [x_1, x_2, ..., x_n]^T \ .$$

The other half-space consists of the points satisfying the inequality

$$\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n \leq \beta$$

i.e.

$$H_- = \{x \in \mathbb{E}^n \,|\alpha^T x \leq \beta\} .$$

The half-space H_+ is called the positive half-space, and the half-space H_- is called the negative half-space.

Let $Q(a_1, a_2, ..., a_n)$ be an arbitrary point of the hyperplane (2.3). Thus

$$\alpha_1 a_1 + \alpha_2 a_2 + ... + \alpha_n a_n = \beta ,$$

i.e.

$$\alpha^T a - \beta = 0 . \tag{2.4}$$

where $a = [a_1, a_2, ..., a_n]^T$. Denote by $M(x_1, x_2, ..., x_n)$ an arbitrary point of $\mathbb{E}^n$, and consider the expression $\alpha^T x - \beta$. By virtue of (2.4) it is possible to write

$$\begin{aligned} \alpha^T x - \beta &= \alpha^T x - \beta - (\alpha^T a - \beta) \\ &= \alpha^T (x - a) \\ &= \alpha_1 (x_1 - a_1) + \alpha_2 (x_2 - a_2) + ... + \alpha_n (x_n - a_n) . \end{aligned} \tag{2.5}$$

The numbers $(x_i - a_i)$ $i = 1, ..., n$, are the components of the vector $\overrightarrow{QM}$. Denote by $\mathbf{n}$ the vector with the components α_i, $i=1, ..., n$. The hyperplane (2.3) consists of the points $M(x_1, ..., x_n)$ for which $<\mathbf{n}, \overrightarrow{QM}> = 0$.

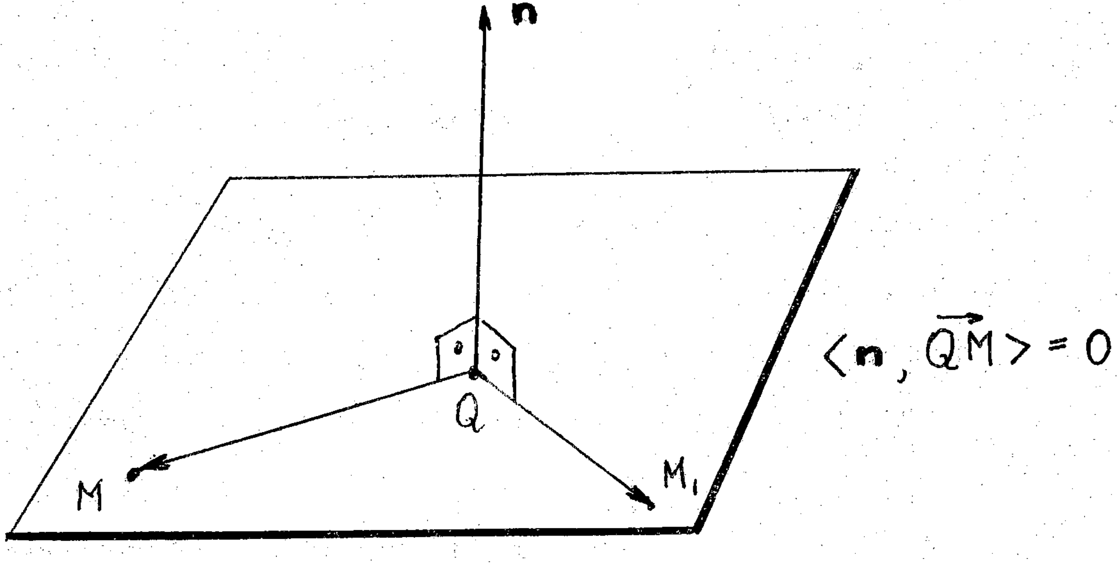


Fig. 2.5. Hyperplane $H = \{x \in E^n \mid \alpha^T x = \beta\}$

In other words, the hyperplane (2.3) consists of the points M for which the vectors **n** and $\overrightarrow{QM}$ are orthogonal. We call the vector **n** the normal to the hyperplane (2.3). Respectively H_+ consists of those points $M(x_1, x_2, \ldots, x_n)$ for which $<\mathbf{n}, \overrightarrow{QM}> \geq 0$, and H_- consists of those points M for which $<\mathbf{n}, \overrightarrow{QM}> \leq 0$.

CONVEX SETS

A set M in the space $\mathbb{E}^n$ is *convex* if it contains line segments that joint each of its points. A line segment between $u, v \in M$, is the set $\{w \in M \mid w = \lambda u + (1 - \lambda) v, \lambda \in [0,1]\}$.

A set M in $\mathbb{E}^n$ is called a *cone* with vertex at the point Q if, together with every point A in M different from Q, the set M contains also the entire ray emanating from Q and passing through A. If the set M is convex and it is a cone then it is called a convex cone.

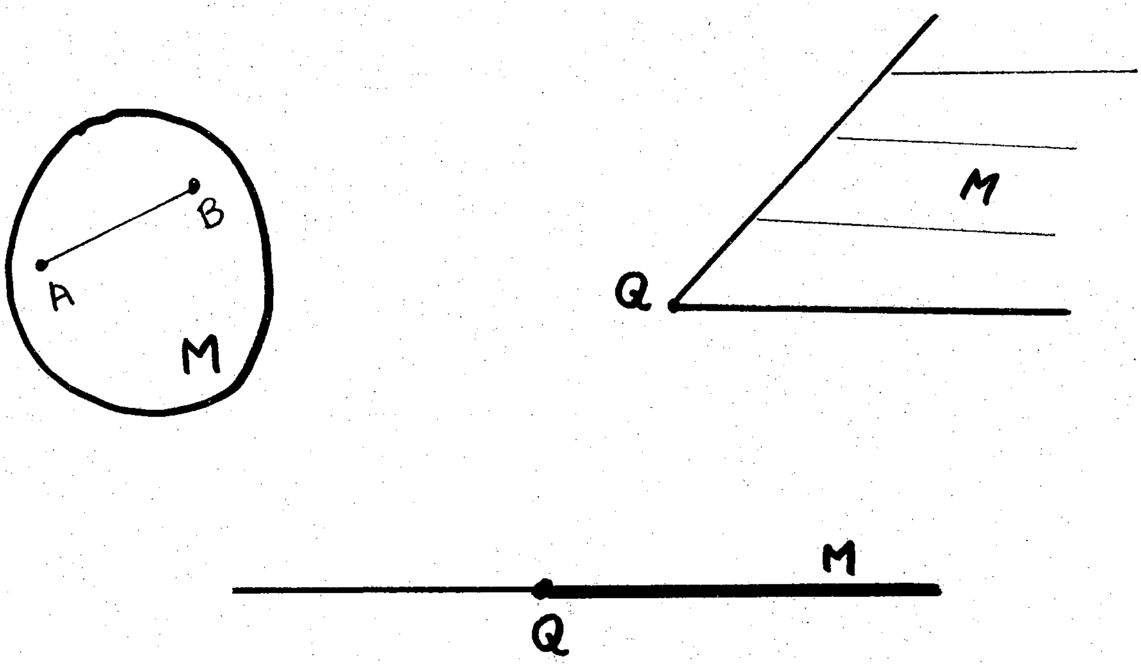


Fig. 2.6. Convex sets

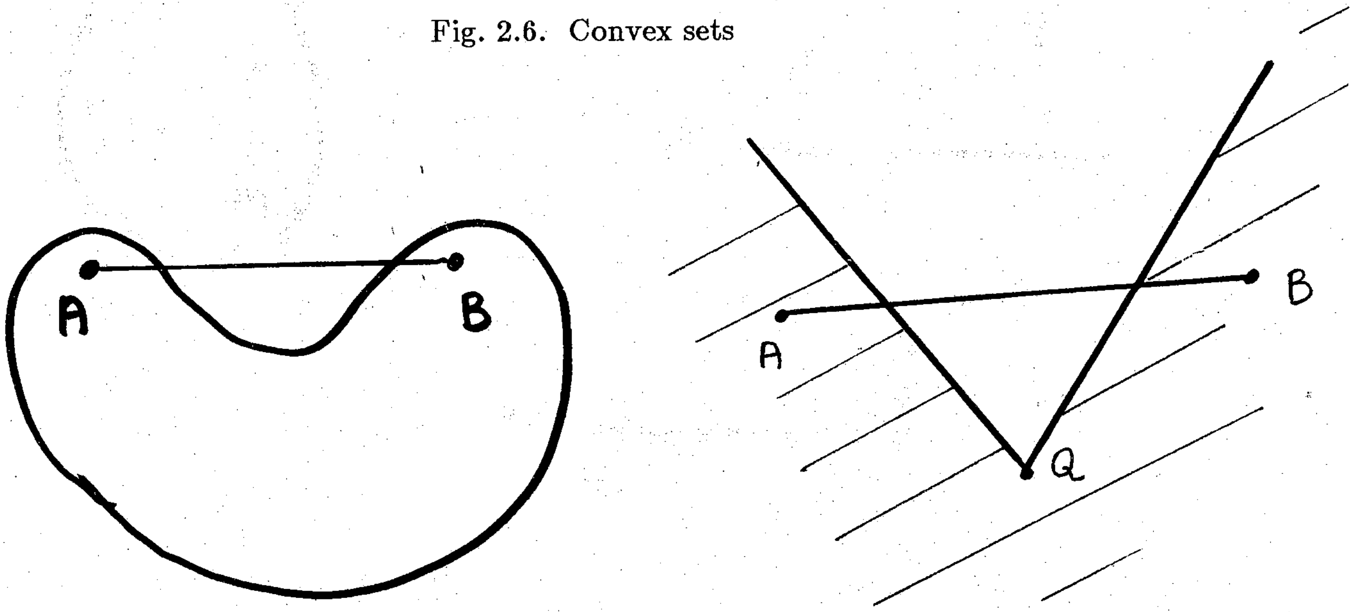


Fig. 2.7. Examples of sets that are not conves

Every convex cone either coincides with the entire space $\mathbb{E}^{n}$ or lies completely in a half-space.

NEIGHBORHOODS

A *neighborhood* of a point $x_0 \in \mathbb{E}^n$ is a set of the form

$$\{x \in \mathbb{E}^n \mid ||x - x_0|| < \epsilon\}$$

where ϵ is some number greater than zero.

The set of all vectors x that satisfy the inequality $||x - x_0|| < \epsilon$ is called an ϵ - *ball* with radius ϵ and center x_0. In the plane, a neighborhood of $x_0 = [x_{10}, x_{20}]^T$ consists of all the points inside of a disc centered at x_0. In three-space, a neighborhood of $x_0 = [x_{10}, x_{20}, x_{30}]^T$ consists of all the points inside of a ball centered at x_0

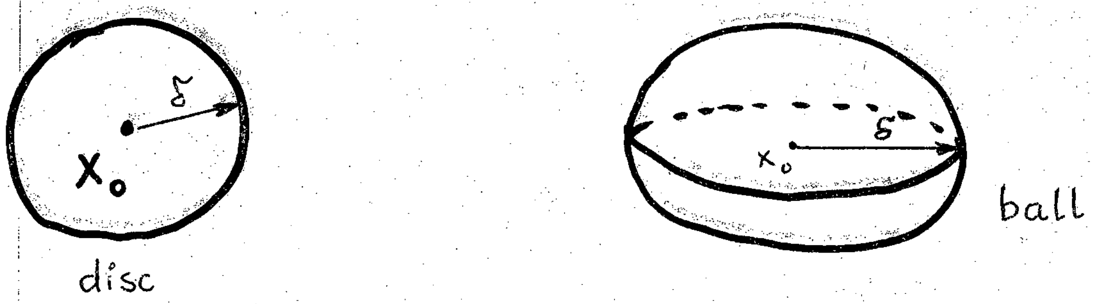


Fig. 2.8. Examples of a neighborhood of a point in $\mathbb{E}^2$ and $\mathbb{E}^3$.

A point x_0 is said to be an *interior point* of the set S if and only if the set S contains some neighborhood of x_0, i.e., if all points within some fixed neighborhood of x_0 are also in S.

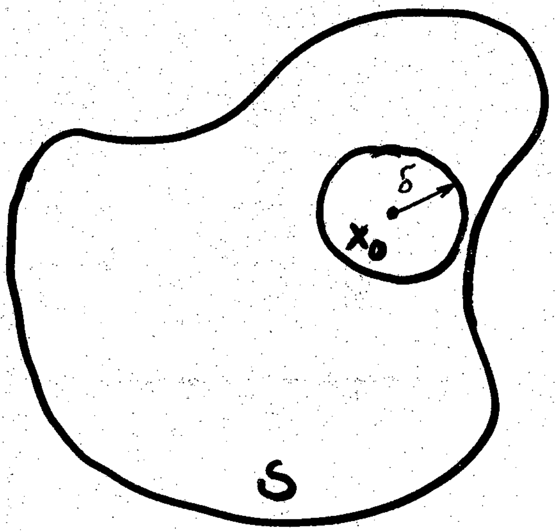


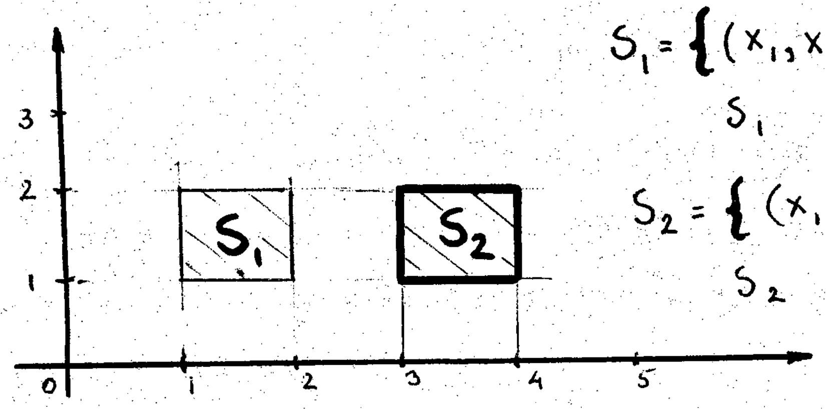
Fig. 2.9. Interior point.

The set of all the interior points of S is called the *interior* of S.

A point x_0 is said to be a ***boundary point*** of the set S if every neighborhood of x_0 contains points that are in S and points that are not in S. The set of all boundary points of S is called the ***boundary*** of S.

A set S is said to be *open* if it contains a neighborhood of each of its points i.e. if each of its points is an interior point or if it contains no boundary points.

A set S is said to be *closed* if it contains its boundary.



$S_1 = \{(x_1, x_2) \mid 1 < x_1 < 2,\ 1 < x_2 < 2\}$

S_1 is open

$S_2 = \{(x_1, x_2) \mid 3 \leq x_1 \leq 4,\ 1 \leq x_2 \leq 2\}$

S_2 is closed

Fig. 2.10. Open and closed sets.

HYPERPLANE OF SUPPORT

Let M be a convex set of the space $\mathbb{E}^n$ and let Q be one of its boundary points. A hyperplane passing through Q is called a *hyperplane of support* of the set M, if the entire set M lies completely in one of the two half-spaces into which this hyperplane divides the space $\mathbb{E}^n$.

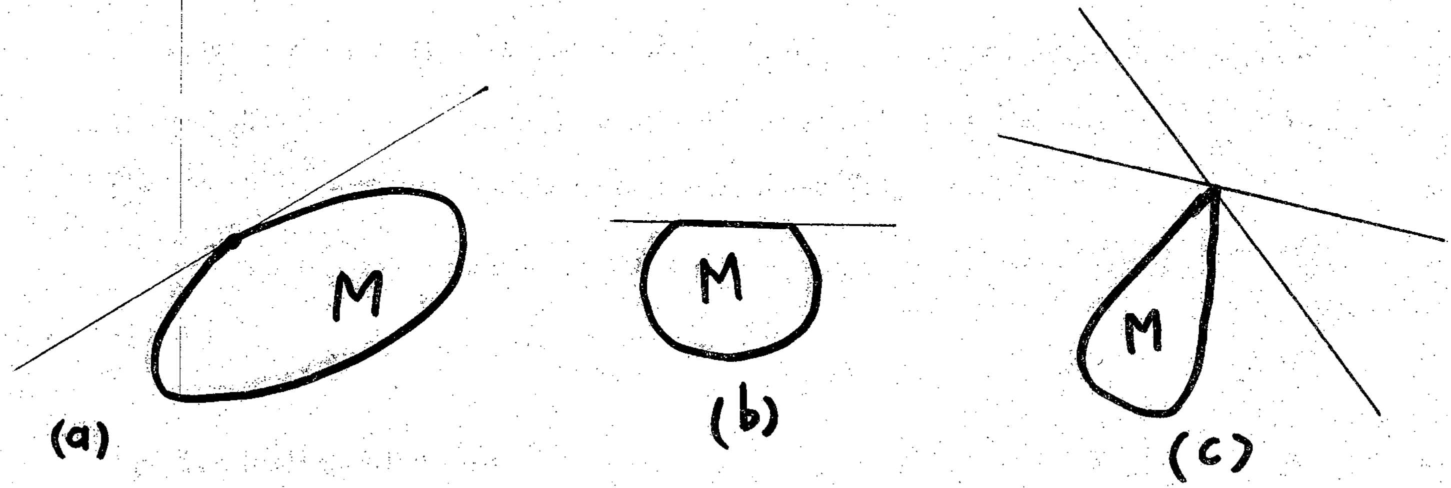


Fig. 2.11. Hyperplane of support for various convex sets (n = 2).

It can be proven that it is possible to draw a hyperplane of support to a convex set M through each of its boundary points. In some cases more than one hyperplane of support can be drawn through some boundary points (Fig. 12 (c)).

In order to show that one can draw a hyperplane of support to the convex set M we need.

The Separation Theorem

Let $M \subset \mathbb{E}^n$ be a closed convex set and let $x_0 \notin M$. Then there exists a vector $\mathbf{a} \in \mathbb{E}^n$ such that

$$\mathbf{a}^T(x - x_0) > 0 \quad \text{for all } x \in M .$$

Proof ([3], [7]):

Let

$$\delta = \min_{x \in M} ||x - x_0|| \ .$$

There is a point z_0 on the boundary of M such that $||z_0 - x_0|| = \delta$. This follows from the theorem of Weierstrass, that is, the continuous function $f = ||x - x_0||$ achieves its minimum over any closed and bounded (compact) set. We consider x in the intersection of the closure of M and the sphere $S_{2\delta}(x_0)$. We claim that $\mathbf{a} = z_0 - x_0$ satisfies the condition of the theorem. (See Fig. 2.12.)

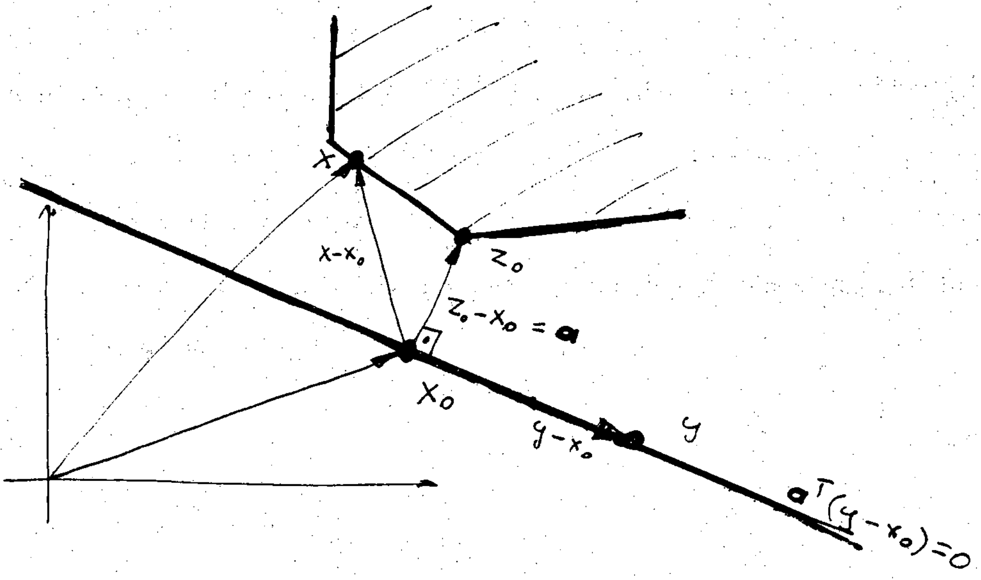


Fig. 2.12. Two dimensional illustration of the Separation Theorem.

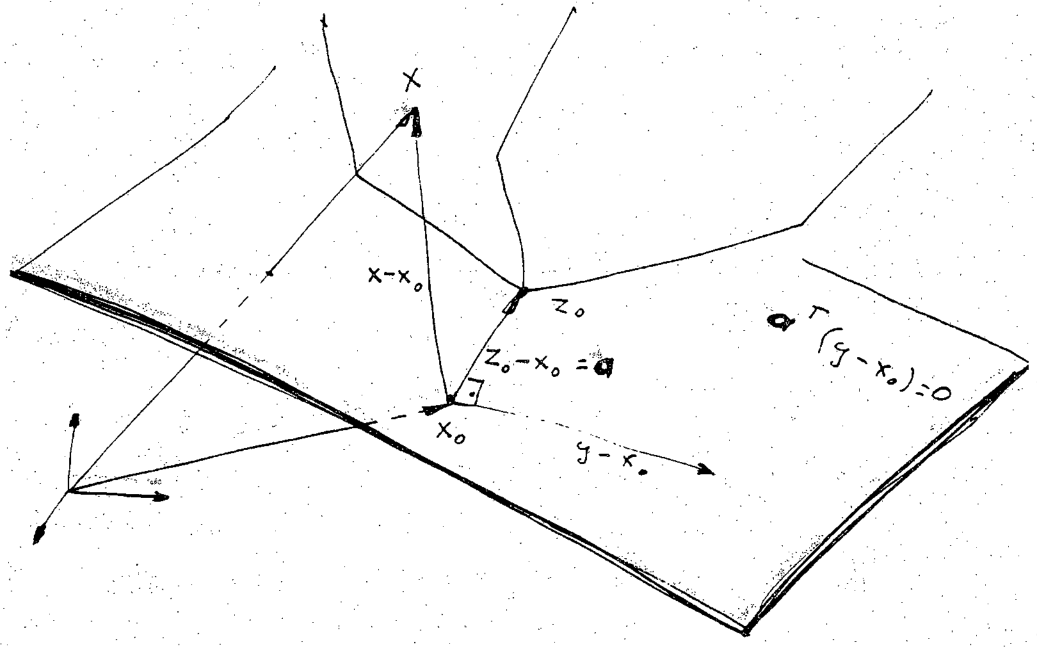


Fig. 2.13. Three dimensional illustration of the Separation Theorem.

Let $x \in M$, then for all $\alpha \in [0,1]$ the point

$$x_\alpha = \alpha x + (1-\alpha) z_0 = z_0 + \alpha(x - z_0) \in M .$$

Hence

$$||x_\alpha - x_0||^2 = ||z_0 + \alpha(x - z_0) - x_0||^2 \geq \ ||z_0 - x_0||^2 .$$

Expansion of the above inequality yields

$$||z_0 - x_0 + \alpha(x - z_0)||^2 = [(z_0 - x_0)^T + \alpha(x - z_0)^T][(z_0 - x_0) + \alpha(x - z_0)]$$

$$= ||z_0 - x_0||^2 + 2\alpha(z_0 - x_0)^T(x - z_0) + \alpha^2||x - z_0||^2 \geq \ z_0 - x_0||^2 .$$

Thus

$$2\alpha(z_0 - x_0)^T(x - z_0) + \alpha^2||x - z_0||^2 \geq \ 0 .$$

Letting $\alpha \longrightarrow 0$ we obtain

$$(z_0 - x_0)^T(x - z_0) \geq 0 .$$

Therefore

$$(z_0 - x_0)^T x \geq (z_0 - x_0)^T z_0 = (z_0 - x_0)^T x_0 + (z_0 - x_0)^T(z_0 - x_0)$$

$$= (z_0 - x_0)^T x_0 + \delta^2 ,$$

and thus

$$(z_0 - x_0)^T(x - x_0) = \mathbf{a}^T(x - x_0) > 0 \qquad \text{for all } x \in M .$$

□

Let K be a convex cone of the space $\mathbb{E}^n$ with vertex Q. If this cone does not coincide with the entire space, then there exists a point A which does not belong to this cone. Therefore none of the points of the ray emanating from Q and passing through A belong to the cone (see Fig. 2.14).

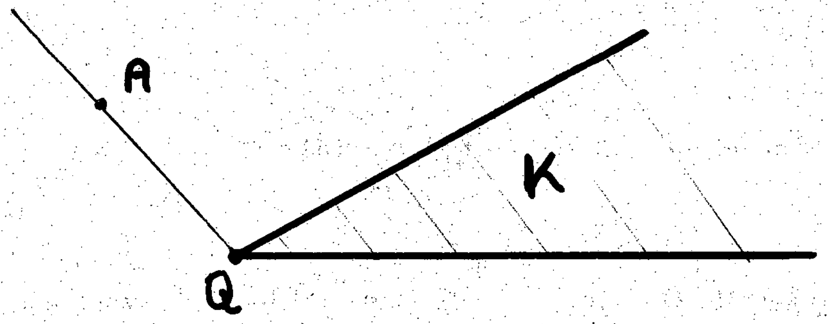


Fig. 2.14. Convex cone K with a ray emanating from its vertex.

Thus, there are points as close as desired to Q which do not belong to K. Hence Q is a boundary point of the convex cone K.

If the convex cone does not coincide with the entire space, then its vertex point is a boundary point of this cone and thus it is possible to draw a hyperplane of support H through Q.

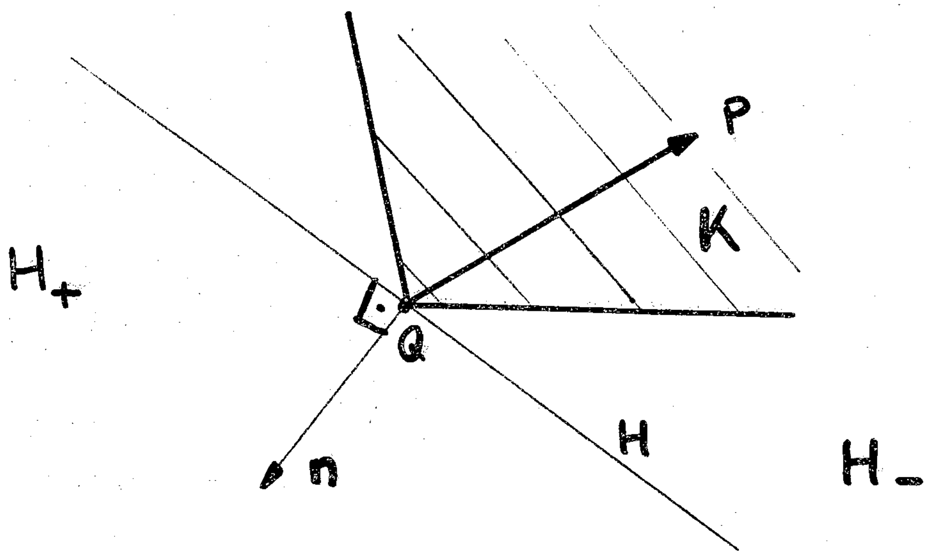


Fig. 2.15. Hyperplane of support H passing through the vertex Q of the convex cone K.

Assume that the cone K lies in the negative half-space H_-. Denote by n the vector that is normal to the hyperplane H. Then we see that if the cone K does not coincide with the entire space $\mathbb{E}^n$, then there exists a non zero vector n such that

$$<\mathrm{n}, \overrightarrow{\mathrm{QP}}> \leq 0$$

for any point P of the cone K.

POLYTOPES AND POLYHEDRA

A set that can be contained in a sphere (ball) of finite radius is called a *bounded set.*

A set is *compact* if it is both closed and bounded.

If M_1 and M_2 are two convex sets, then their intersection is also a convex set (provided that this intersection contains at least one point.)

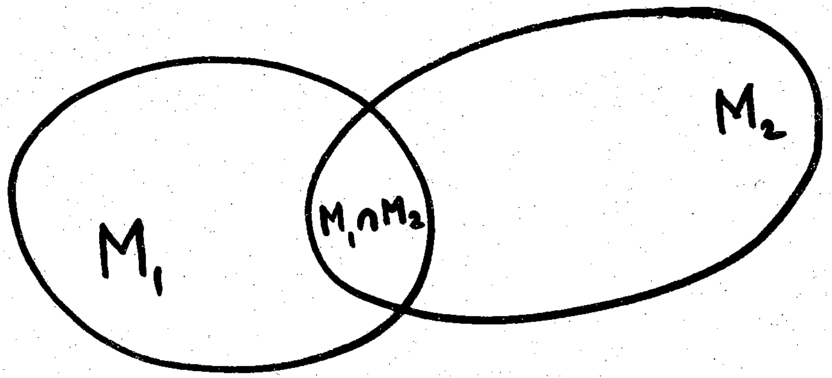


Fig. 2.16. Intersection of two convex sets.

In fact, the intersection of any number of convex sets is convex. In what follows we will be concerned with the intersection of a finite number of half-spaces. Since every half space $\alpha^T x \leq \beta$ (or $\alpha^T x \geq \beta$) is convex in $\mathbb{E}^n$, the intersection of any number of half spaces is a convex set.

A set which can be expressed as the intersection of a finite number of closed half spaces is said to be a *convex polytope*.

A nonempty bounded polytope is called a *polyhedron*.

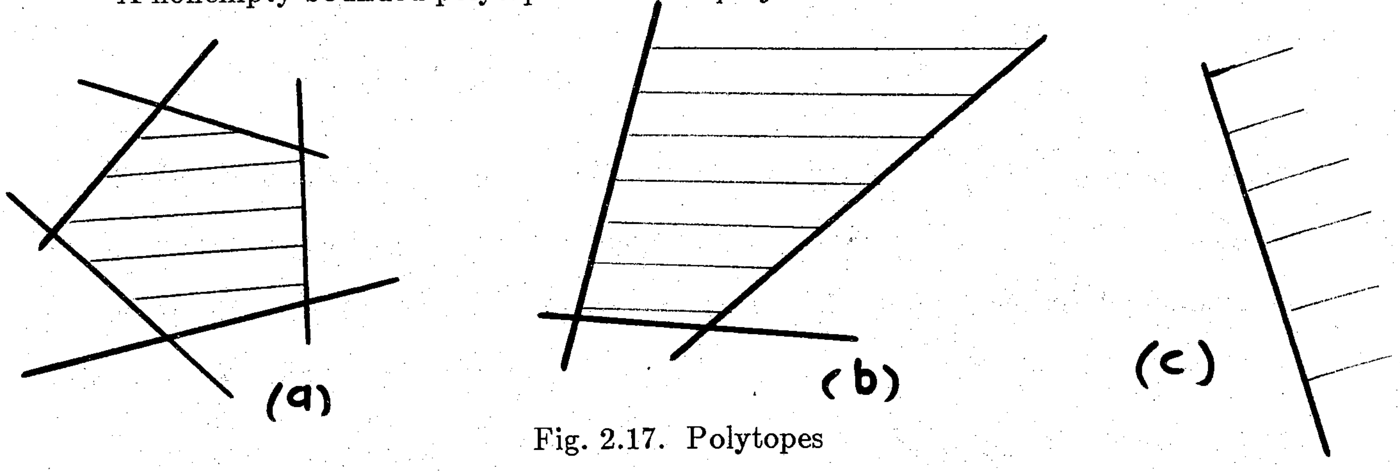


Fig. 2.17. Polytopes

For every convex polyhedron M in an n-dimensional space there exists a non-negative integer such that M is contained in a k-dimensional plane of the n-dimensional space, but is not entirely contained in any (k-1)-dimensional plane. Furthermore, there exists only one k-dimensional plane containing M. It is called the *carrier* of the polyhedron M, and k is called the dimension of M. A zero-dimensional polyhedron is a point of n-dimensional space. A one-dimensional polyhedron is a segment, and its carrier is the straight line on which it lies. The boundary of any k-dimensional polyhedron (k>0) consists of a finite number of k-1 dimensional polyhedra [2]. For example, the boundary of a one-dimensional polyhedron consists of two points which are the endpoints of the segment.

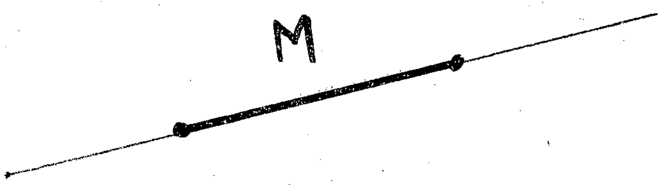


Fig. 2.18. One-dimensional polyhedron.

These (k-1)-dimensional polyhedra are called the (k-1)-dimensional *faces* of the k-dimensional polyhedron. Each of these faces has in turn (k-2)-dimensional faces. Thus, every k-dimensional polyhedron has faces of dimensions k-1, k–2,..., 1, 0. The zero-dimensional faces of a polyhedron are called its *vertices,* and the one-dimensional faces are called *edges.*

EXTREME POINTS

A point x in a convex set M is said to be an *extreme point* of M if there are no two distinct points x_1 and x_2 in M such that $x = \lambda x_1 + (1 - \lambda)x_2$ for some λ, $0 < \lambda < 1$.

The *convex hull* of a set S, denoted co(S) is the set which is the intersection of all convex sets containing S. This intersection is the smallest convex set containing S. For example, the convex hull of three points not lying on a straight line is a triangle

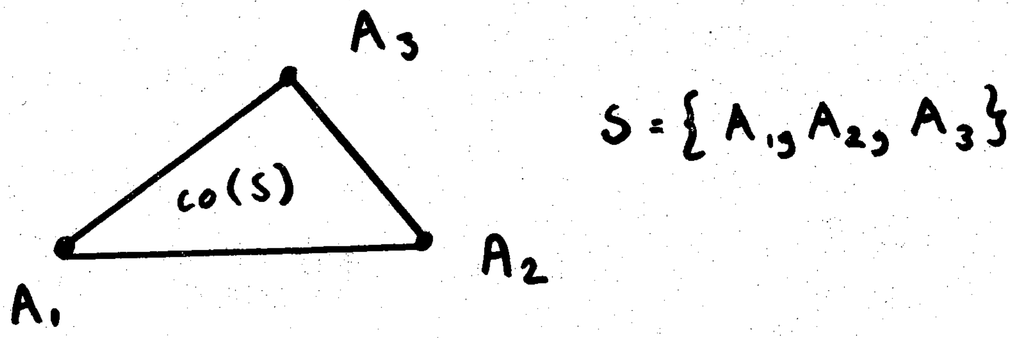


Fig. 2.19. Convex hull of the set which consists of three points.

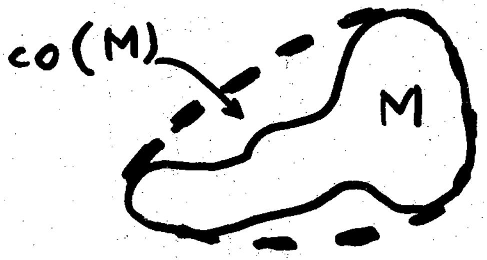


Fig. 2.20. Convex hull of the set M.

Any convex polyhedron is the convex hull of all its vertices.

In general, a closed bounded convex set in $\mathbb{E}^n$ is equal to the closed hull of its extreme points.

3. LINEAR TRANSFORMATIONS AND MATRICES

A correspondence A that associates each point in a space X with a point in a space Y is said to be a *mapping* or *transformation* from X to Y. For convenience this situation is symbolized by

$$A\colon\ X \to Y$$

A transformation A is linear if it satisfies the following properties

1) $A(ax) = a(Ax)$
2) $A(x+y) = Ax + Ay$

One can obtain concrete realizations of such linear transformations in terms of matrices. Suppose $\{e_1, e_2, \dots, e_n\}$ is a basis. Application of the transformation A to each vector from the basis results in n new vectors

$$e_i' = A\, e_i, \quad i = 1,2,\dots,n.$$

Since the e_i form a basis, it is possible to express e_i' in terms of the e_i as follows

$$A\, e_i = e_i' = \sum_{j=1}^{n} a_{ji}\, e_j, \quad i = 1,2,\dots,n.$$

The n^2 coefficients a_{ij} determine the linear transformation. The coordinates of any vector of the space are changed according to the formula

$$x' = A\, x$$

where

$$x' = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Let us now consider a linear transformation A, and let the matrix A correspond to it with respect to the basis $\{e_1,e_2,...,e_n\}$, and the matrix B with respect to the basis $\{e_1',e_2',...,e_n'\}$. Let x be the column of the coordinates of some vector with respect to the basis $\{e_1,e_2, . . . ,e_n\}$ and x' coordinates of the same vector with respect to the basic $\{e_1',e_2',...,e_n'\}$. Denote by T the transition matrix from e_i to e_i'. We have

$$y = Ax,$$
$$y' = Bx'.$$

But $x' = Tx$ and $y' = Ty$. Therefore

$$y' = Ty = TAx = Bx' = BTx,$$

and

$$A = T^{-1}BT.$$

In conclusion: similar matrices correspond to the same linear transformation with respect to different basis.

In many applications we will utilize the concept of the Hermitian transpose A^* defined by

$$A^* = \overline{A}^T$$

where the bar in $\overline{A}^T$ denotes the complex conjugate of a transposed matrix A^T. Thus A^* is obtained by replacing every component of the A^T by its complex conjugate. In the case where A is a real matrix $A^* = A^T$. If $A^* = A$, we may say that the matrix A is hermitian. A real hermitian matrix is called symmetric. A simple calculation reveals that

$$< Ax,y > = < x, A^*y > .$$

EIGENVALUES AND EIGENVECTORS

Corresponding to an nxn square matrix A, a scalar λ and a nonzero vector v satisfying the equation $Av=\lambda v$ are said to be, respectively, an eigenvalue and eigenvector of A. In order that λ be an eigenvalue it is necessary and sufficient condition for the matrix $\lambda I - A$ to be singular, that is $\det[\lambda I - A]=0$. This leads to an nth-order polynomial equation

$$\det[\lambda I - A] = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0.$$

This equation must, according to the fundamental theorem of algebra, have n (possibly nondistinct) complex roots which are the eigenvalues of A. Suppose now that the determinant equation $\det[\lambda I - A] = 0$ has n *real distinct* roots: $\lambda_1, \lambda_2, \ldots, \lambda_n$. Corresponding to them we find n vectors $v_1, v_2, \ldots, v_n$ such that

$$Av_i = \lambda_i v_i \ , \quad i = 1,2,...,n \ .$$

We shall show that the eigenvectors $v_1, v_2, ..., v_n$ are linearly independent. We prove the above statement by contradiction. In order to do this assume that there exists a set of scalars c_i such that

$$\sum_{i=1}^{n} c_i \, v_i = 0 \ ,$$

where at least one of the c_i does not vanish. Without loss of generality suppose that $c_1 \neq 0$. Next, consider the matrix

$$P = (\lambda_2 I - A)(\lambda_3 I - A) \cdot \cdots \cdot (\lambda_n I - A) \ .$$

Note that

$$Pv_n = (\lambda_2 I - A)(\lambda_3 I - A) \cdot \cdots \cdot (\lambda_{n-1} I - A)(\lambda_n I - A)v_n$$

$$= (\lambda_2 I - A)(\lambda_3 I - A) \cdot \cdots \cdot (\lambda_{n-1} I - A)(\lambda_n v_n - Av_n)$$

$$= 0 \qquad \text{because} \qquad (\lambda_n v_n - Av_n) = 0.$$

Similarly

$$Pv_k = 0, \qquad k = 2,3,\ldots,n.$$

But

$$Pv_1 = (\lambda_2 I - A)(\lambda_3 I - A) \cdot \cdots \cdot (\lambda_{n-1} I - A)(\lambda_n I - A)v_1$$

$$= (\lambda_2 I - A)(\lambda_3 I - A) \cdot \cdots \cdot (\lambda_{n-1} v_1 - Av_1)(\lambda_n - \lambda_1)$$

$$= (\lambda_2 I - A)(\lambda_3 I - A)v_1 \cdot \cdots \cdot (\lambda_{n-1} - \lambda_1)(\lambda_n - \lambda_1)$$

$$= (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1) \cdot \cdots \cdot (\lambda_{n-1} - \lambda_1)(\lambda_n - \lambda_1)v_1.$$

Using the above equation we see that

$$P(\sum_{i=1}^{n} c_i v_i) = \sum_{i=1}^{n} c_i \, Pv_i = c_1 \, Pv_1$$

$$= c_1(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1 \cdot \cdots \cdot (\lambda_n - \lambda_1)v_1 = 0.$$

Since all λ_i are distinct, it must follow that $c_1 = 0$. It thus follows that all c_i must vanish, and therefore the set of eigenvectors $\{v_1, v_2, \ldots, v_n\}$ is linearly independent.

□

Linearly independent set of eigenvectors can be considered as a basis. With respect to this basis matrix A has a diagonal from. Indeed, let

$$T = [v_1, \; v_2, \ldots, v_n]$$

Then

$$T^{-1}AT = T^{-1}A[v_1, v_2, \ldots, v_n]$$

$$= T^{-1}[Av_1, Av_2, \ldots, Av_n]$$

$$= T^{-1}[\lambda_1 v_1, \lambda_2 v_{2,\ldots,} \lambda_n v_n]$$

$$= \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix}.$$

Let us now consider hermitian matrices. We shall show that all eigenvalues of a hermitian matrix are real [4]. Let

$$Ax = \lambda x ,$$

where $x \neq 0$. Taking the inner product of Ax with x yields

$$< Ax,x > = < \lambda x,x > = \lambda < x,x > .$$

On the other hand

$$< Ax,x > = < x,A^* x > = < x,Ax > = < x,\lambda x > = \bar{\lambda} < x,x > .$$

We note that $< x,x >$ is real and $< x,x > > 0$. Hence

$$\lambda < x,x > = \bar{\lambda} < x,x >$$

and

$$(\lambda - \bar{\lambda}) < x,x > = 0 .$$

But since $< x,x > > 0$, we have

$$\lambda = \bar{\lambda}$$

so λ is real.

□

We shall now show that the eigenvectors associated with distinct eigenvalues of a hermitan matrix are orthogonal. Suppose

$$Av_1 = \lambda_1 v_1 ,$$

$$Av_2 = \lambda_2 v_2$$

where

$$\lambda_1 \neq \lambda_2.$$

Then

$$< Av_1, v_2 > = < \lambda_1 v_1, v_2 > = \lambda_1 < v_1, v_2 >,$$

but since $A = A^*$

$$< Av_1, v_2 > = < v_1, A^* v_2 > = < v_1, Av_2 >$$

$$= \lambda_2 < v_1, v_2 >.$$

Therefore

$$\lambda_1 < v_1, v_2 > = \lambda_2 < v_1, v_2 >,$$

and since $\lambda_1 \neq \lambda_2$, it follows that

$$< v_1, v_2 > = 0.$$

If A is hermitian and its eigenvalues are distinct, then the set of its eigenvectors forms an orthogonal basis for $\mathbb{E}^n$. If the basis $\{v_1, v_2, ..., v_n\}$ is normalized so that each element has norm unity, then defining the matrix

$$Q = [v_1, v_2, ..., v_n] ,$$

we have

$$Q^T Q = I,$$

and hence

$$Q^T = Q^{-1} .$$

A matrix with this property is said to be an *orthogonal* matrix

A symmetric matrix $\mathbf{A}$ is said to be *positive definite* if the *quadratic form* $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is positive for all nonzero vectors $\mathbf{x}$. Similarly, we define *positive semidefinite, negative definite,* and *negative semidefinite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$, <0, or ≤ 0 for x. The matrix $\mathbf{A}$ is *indefinite if* $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is positive for some $\mathbf{x}$ and negative for others.

It is easy to obtain a connection between definiteness and the eigenvalues of $\mathbf{A}$. For any $\mathbf{x}$ let $\mathbf{y} = \mathbf{Q}^{-1}\mathbf{x}$ where $\mathbf{Q}$ is defined as above. Then $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{Q}^T\mathbf{A}\mathbf{Q}\mathbf{y} = \sum_{i=1}^{n} \lambda_i y_i^2$. Since the y_i's are arbitrary (since x is), is is clear that $\mathbf{A}$ is positive definite (or positive semidefinite) if and only if all eigenvalues of $\mathbf{A}$ are positive (or nonnegative).

Through diagonalization we can show that a positive semidefinite matrix $\mathbf{A}$ has a positive semidefinite (symmetric) square root $\mathbf{A}^{1/2}$ satisfying $\mathbf{A}^{1/2} \cdot \mathbf{A}^{1/2} = \mathbf{A}$ [7]. For this we use $\mathbf{Q}$ as above and define

$$\mathbf{A}^{1/2} = \mathbf{Q} \begin{bmatrix} \lambda_1^{1/2} & & & & \\ & \lambda_2^{1/2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \lambda_n^{1/2} \end{bmatrix} \mathbf{Q}^T,$$

which is easily verified to have the desired properties.

NORMS OF MATRICES

The norm of a matrix A is a nonnegative number $\|A\|$ satisfying the conditions

1) $\|A\| > 0$ if $A \neq 0$ and $\|0\| = 0$;

2) $\|cA\| = |c| \, \|A\|p=;$

3) $\|A+B\| \leq \|A\| + \|B\|$

4) $\|AB\| \leq \|A\| \, \|B\|$.

The norm of a matrix may be chosen in a variety of ways. In many problems both matrices and vectors appear simultaneously. Therefore, it is convenient to introduce the norm of a matrix in such a way that it will be connected with the vector norms employed in the considerations.

We shall say that the norm of a matrix is *induced* or it is *compatible* ([5]) with a given norm of vectors if for any matrix A and any vector x the following inequality is satisfied

$$\|Ax\| \leq \|A\| \, \|x\|.$$

We define the norm of the matrix A as the maximum of the norms of the vectors Ax where the vector x runs over the set of all vectors whose norm equals unity:

$$\|X\| = \max_{\|x\|=1} \|Ax\|.$$

Because of the continuity of a norm, for each matrix A this maximum is attainable, that is, a vector x_0 can be found such that $\|x_0\| = 1$ and $\|Ax_0\| = \|A\|$. We shall now prove that a norm defined in such a manner satisfies conditions (1) - (4), and the compatibility condition.

Proof [5]:

(1) Let $A \neq 0$. Then a vector x, $\|x\|=1$, can be found such that $Ax \neq 0$, and thus $\|Ax\| \neq 0$. Hence $\|A\| = \max\limits_{\|x\|=1} \|Ax\| \neq 0$. If, on the other hand, $A=0$, $\|A\| = \max\limits_{\|x\|=1} \|0x\| = 0$.

(2) By the definition

$$\|cA\| = \max \|cAx\|.$$

Obviously

$$\|cAx\| = |c| \; \|Ax\|$$

and therefore

$$\|cA\| = \max_{\|x\|=1} |c| \; \|Ax\| = |c| \max_{\|x\|=1} \|Ax\|$$

$$= |c| \; \|A\|.$$

The Compatibility Condition

Let $y \neq 0$ be any vector. Then $x = \frac{1}{\|y\|} y$ will satisfy the condition $\|x\|=1$. Consequently

$$\|Ay\| = \|A(\|y\|x)\| = \|y\| \; \|Ax\| \leq \|y\| \; \|A\|.$$

(3) For the matrix $A+B$ find a vector x_0 such that

$$\|A+B\| = \|(A+B)x_0\| \text{ and } \|x_0\|=1.$$

Then we have

$$\|A+B\| = \|(A+B)x_0\|$$

$$= \|Ax_0 + Bx_0\| \leq \|Ax_0\| + \|Bx_0\|$$

$$\leq \|A\| \|x_0\| + \|B\| \|x_0\| = \|A\| + \|B\| .$$

(4) Fop the matrix AB find a vector x_0 such that $\|x_0\|=1$ and $\|ABx_0\|=\|AB\|$. Then we have

$$\|AB\| = \|ABx_0\|$$

$$= \|A(Bx_0)\| \leq \|A\| \|Bx_0\| \leq \|A\| \|B\| \|x_o\|$$

$$= \|A\| \|B\|.$$

□

Let

$$\|x\| = \left(\sum_{k=1}^{n} x_k^2 \right)^{\frac{1}{2}} = \sqrt{< x,x >}$$

The induced matrix norm by this vector norm is

$$\|A\| = \sqrt{\lambda_1}$$

where λ_1 is the largest eigenvalue of the matrix $A^* A$.

Proof:

We have

$$\|A\| = \max_{\|x\|=1} \|Ax\|.$$

On the other hand

$$||Ax||^2 = <Ax, Ax> = <x, A^*Ax>$$

The matrix A^*A is a hermitian one. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_n$ be its eigenvalues and $x_1, x_2, ..., x_n$ be the orthonormal set of the eigenvectors corresponding to these eigenvalues.

Now we take an arbitrary vector x with $||x|| = 1$ and represent it as a linear combination of x_i, that is:

$$x = c_1x_1 + c_2x_2 + \cdots + c_nx_n.$$

Note that

$$<x,x> = c_1^2 + c_2^2 + \cdots + c_n^2 = 1.$$

Furthermore

$$||Ax||^2 = <x, A^*Ax>$$

$$= <c_1x_1 + \cdots + c_nx_n, c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n>$$

$$= \lambda_1c_1^2 + \cdots + \lambda_nc_n^2 \leq \lambda_1(c_1^2 + \cdots + c_n^2) = \lambda_1.$$

For the eigenvector x_1 of A^*A corresponding to the eigenvalue λ_1 we have

$$||Ax_1||^2 = <x_1, A^*Ax_1> = <x_1, \lambda_1x_1> = \lambda_1 ,$$

and hence

$$\max_{||x|| = 1} ||Ax|| = \sqrt{\lambda_1} .$$

□

From the above considerations one can deduce the following important inequality. If an n×n matrix P is real symmetric positive definite, then

$$\lambda_{\min}(P) ||x||^2 \leq x^TPx \leq \lambda_{\max}(P) ||x||^2 .$$

We shall refer to the above relation as the Rayleigh inequality.

Examples:

Consider the matrix operator $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ from $\mathbb{R}^2$ into $\mathbb{R}^2$ and let the norm in $\mathbb{R}^2$ be given by

$$||x|| = (x_1^2 + x_2^2)^{1/2} .$$

Then $A^T A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$,

$$\det[\lambda I - A^T A] = \lambda^2 - 10\lambda + 9 = (\lambda - 1)(\lambda - 9) .$$

Thus

$$||A|| = \sqrt{9} = 3 .$$

The eigenvector of $A^T A$ corresponding to $\lambda_1 = 9$ is

$$x_1 = \frac{1}{\sqrt{2}} [1 \;\; 1]^T .$$

Note that

$$||Ax_1|| = ||A|| .$$

Indeed

$$||Ax_1|| = \left\| \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \frac{1}{\sqrt{2}} \left\| \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\|$$

$$= \frac{1}{\sqrt{2}} \sqrt{3^2 + 3^2} = 3 .$$

In this example $||A|| = \max\limits_{1 \le i \le n} |\lambda_i(A)|$. This is because $A = A^T$.

Warning: In general max $\lambda(A) \neq ||A||$, instead we have $||A|| \geq \max_{1\leq i\leq n} |\lambda_i(A)|$. For example, let

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} , \text{ then } A^T A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

and $[\lambda I - A^T A] = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda - 1 \end{bmatrix}$. Hence

$$\det[\lambda I - A^T A] = \lambda(\lambda - 1) .$$

Thus

$$||A|| = 1 .$$

The eigenvector of $A^T A$ corresponding to $\lambda_1 = 1$ is any nonzero vector of

$$\text{Adj}[\lambda I - A^T A]\Big|_{\lambda=1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Take

$$x_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \text{ then } Ax_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

and

$$||Ax_1|| = \left\|\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\| = 1 = ||A|| .$$

The fact that

$$||A|| \geq \max_{1\leq i\leq n} |\lambda_i(A)|$$

can be shown as follows [5]. Consider a matrix

$$B = \frac{1}{||A|| + \epsilon} A ,$$

where ϵ is any positive real number. We have

$$||B|| = \frac{||A||}{||A|| + \epsilon} < 1 .$$

Hence $B^m \longrightarrow 0$ as $m \longrightarrow \infty$, and thus

$$|\lambda_i(B)| < 1 .$$

On the other hand

$$\lambda_i(B) = \frac{1}{||A|| + \epsilon} \lambda_i(A) .$$

Thus

$$|\lambda_i(B)| = \frac{|\lambda_i(A)|}{||A|| + \epsilon} < 1$$

that is

$$|\lambda_i(A)| < ||A|| + \epsilon .$$

Since ϵ may be arbitrarily small

$$|\lambda_i(A)| \leq ||A|| .$$

QUADRATIC FORMS

A real *quadratic form* is an expression

$$z = x^T Ax ,$$

where x is an n×1 real column vector and A is an n×n real symmetric matrix, that is $A = A^T$.

If the matrix A is not a symmetric one we can always replace it with a symmetric one, say $A_o = A_o^T$ such that

$$x^T Ax = x^T A_o x .$$

This is because

$$x^T Ax = x^T \left(\frac{1}{2} A + \frac{1}{2} A^T \right) x$$

Note that

$$A_o = \frac{1}{2} (A + A^T)$$

is a symmetric matrix. Hence there is no loss of generality in supposing A to be symmetric.

A quadratic form $x^T Ax$ is said to be *positive definite* if $x^T Ax > 0$ for all nonzero vectors x. It is *positive semidefinite* if $x^T Ax \geq 0$ for all x. Similarly, we define *negative definite* and *negative semidefinite* if $x^T Ax < 0$ or ≤ 0 for all x.

The minors of a matrix A are the determinants of the matrices obtained by removing successively rows and columns from A. The *leading principal minors* are det A and the minors obtained by removing successively the last row and the last column. The *principal minors* are det A itself and the determinants of matrices obtained by removing successively an ith row and an ith column. Thus the leading principal minors of an n×n matrix A are:

$$\Delta_1 = a_{11} , \quad \Delta_2 = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \Delta_3 = \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \ldots , \Delta_n = \det A .$$

The principal minors are

$$\Delta_p \begin{pmatrix} i_1, i_2, \ldots, i_p \\ i_1, i_2, \ldots, i_p \end{pmatrix}, \quad 1 \le i_1 < i_2 < \ldots < i_p \le n ,$$

$$p = 1, 2, \ldots, n .$$

where

$$\Delta_p \begin{pmatrix} i_1, i_2, \ldots, i_p \\ i_1, i_2, \ldots, i_p \end{pmatrix} = \det \begin{bmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \cdots & a_{i_1 i_p} \\ a_{i_2 i_1} & a_{i_2 i_2} & \cdots & a_{i_2 i_p} \\ \vdots & \vdots & & \vdots \\ a_{i_p i_1} & a_{i_p i_2} & \cdots & a_{i_p i_p} \end{bmatrix} .$$

We will next prove Sylvester's criterion which allows one to determine if a quadratic form is positive definite.

Sylvester's Criterion

A quadratic form $x^T Ax$ $(A = A^T)$ is positive definite if and only if its leading principal minors are positive.

Note that this criterion does not hold if A has not been symmetrized. To see this consider an example where

$$A = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} .$$

The leading principal minors of A are

$$\Delta_1 = 1 > 0 \quad \text{and} \quad \Delta_2 = \det A = 1 > 0 \,.$$

On the other hand

$$x^T A x = x^T \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} x = \frac{1}{2} x^T \left(\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \right) x$$

$$= x^T \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} x = x^T A_o x \,.$$

The leading principal minors of A_o are

$$\Delta_1 = 1 > 0 \quad \text{and} \quad \Delta_2 = \det A_o = -3 < 0 \,.$$

From Sylvester's criterion it follows that a necessary condition for a real quadratic form to be positive semidefinite is that the leading principal minors be nonnegative. However, this is *NOT* a sufficient condition.

A real quadratic form is positive semidefinite if and only if all principal minors are nonnegative.

The key to the proof of Sylvester's criterion is the fact that a quadratic form can be expressed in some basis as a sum of squares

$$z = \frac{\Delta_0}{\Delta_1} \bar{x}_1^2 + \frac{\Delta_1}{\Delta_2} \bar{x}_2^2 + \dots + \frac{\Delta_{n-1}}{\Delta_n} \bar{x}_n^2 \,,$$

where $\Delta_0 \triangleq 1$, and $\bar{x}_i$ are the coördinates of the vector x in a new basis.

REDUCTION OF A QUADRATIC FORM INTO ITS CANONICAL REPRESENTATION

We shall now describe a method of constructing a basis in which a given real quadratic form becomes a sum of squares [4].

Consider a quadratic form $x^T Ax$, where $A = A^T$. Using the inner product notation one can represent $x^T Ax$ as

$$x^T Ax = <x,Ax> = <Ax,x> .$$

Note that

$$<x_1 + x_2,Ax> = <x_1,Ax> + <x_2,Ax> ,$$

and

$$<\alpha x,Ay> = \alpha<x,Ay> .$$

Now let $f_1,f_2,...,f_n$ be a basis for $\mathbb{R}^n$, and let

$$x = x_1 f_1 + x_2 f_2 + ... + x_n f_n .$$

We shall express the quadratic form using the coordinates $x_i (i = 1,...,n)$ of x relative to the basis $f_1,f_2,...,f_n$. We have

$$z = x^T Ax = < x_1 f_1 + x_2 f_2 + ,..., + x_n f_n, A(x_1 f_1 + x_2 f_2 + ... + x_n f_n)>$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j <f_i,Af_j> .$$

If we denote the constants

$$<f_i,Af_j> = a_{ij}$$

then

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = x^T A x .$$

The matrix A is called the matrix of the quadratic form z relative to the basis $f_1, f_2, \dots, f_n$. If we change the basis from $f_1, f_2, \dots, f_n$ to $f_1', f_2', \dots, f_n'$ then the coordinates of the vector x in the new basis are expressed in terms of the coordinates in the old basis as

$$x = [f_1', f_2', \dots, f_n']\, \bar{x} = F\bar{x} .$$

Accordingly the matrix of the quadratic form z in the new basis is

$$z = x^T A x = \bar{x}^T F^T A F \bar{x} .$$

Now, let the quadratic form z be defined relative to the basis $e_1, e_2, \dots, e_n$ as

$$z = x^T A x$$

where $a_{ij} = \langle e_i, Ae_j \rangle$. Our goal is to find a new basis, say, $f_1, f_2, \dots, f_n$ such that the matrix of the quadratic form in the new basis is a diagonal one, that is

$$\langle f_i, Af_j \rangle = 0 \quad \text{for} \quad i \neq j .$$

We shall seek the new basis in the form

$$\begin{aligned} f_1 &= \alpha_{11} e_1 \\ f_2 &= \alpha_{21} e_1 + \alpha_{22} e_2 \\ &\vdots \\ f_n &= \alpha_{n1} e_1 + \alpha_{n2} e_2 + \dots + \alpha_{nn} e_n . \end{aligned}$$

Observe that if

$$\langle f_i, Ae_j \rangle = 0 \quad \text{for} \quad j = 1,2,\dots,i-1 ,$$

then

$$\langle f_i, Af_j \rangle = 0 \quad \text{for} \quad j = 1,2,\dots,i-1 .$$

Indeed

$$<f_i, A(\alpha_{j1}e_1 + \alpha_{j2}e_2 + ... + \alpha_{jj}e_j)> = 0$$

for $j = 1,2,...,i-1$.

Our goal then is to determine the coefficients $\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jj}$ such that the vector

$$f_j = \alpha_{j1}e_1 + \alpha_{j2}e_2 + ... + \alpha_{jj}e_j$$

satisfies the following relations

$$<e_i, Af_j> = 0 \quad \text{for} \quad i = 1,2,...,j-1$$

and

$$<e_j, Af_j> = 1 .$$

The above j relations determine the vector f_j in a unique way. Indeed, upon substituting the expression for f_j into the above equations we obtain a set of the equations of the form

$$\begin{array}{llll} \alpha_{j1}<e_1,Ae_1> & + \alpha_{j2}<e_1,Ae_2> & + ... & + \alpha_{jj}<e_1,Ae_j> = 0 \\ \vdots & \vdots & & \vdots \\ \alpha_{j1}<e_{j-1},Ae_1> & + \alpha_{j2}<e_{j-1},Ae_2> & + ... & + \alpha_{jj}<e_{j-1},Ae_j> = 0 \\ \alpha_{j1}<e_j,Ae_1> & + \alpha_{j2}<e_j,Ae_2> & + ... & + \alpha_{jj}<e_j,Ae_j> = 1 . \end{array}$$

The above set of equations is equivalent to the following matrix equation

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & & & \\ a_{j1} & a_{j2} & \cdots & a_{jj} \end{bmatrix} \begin{bmatrix} \alpha_{j1} \\ \alpha_{j2} \\ \vdots \\ \alpha_{jj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} .$$

If the leading principal minors of the matrix A are not equal to zero then the coefficients α_{ji} can be obtained by employing Cramer's rule. In particular

$$\alpha_{jj} = \frac{\Delta_{j-1}}{\Delta_j}$$

and

$$\alpha_{ji} = \frac{\det \begin{bmatrix} a_{11} & \cdots & a_{1i-1} & 0 & a_{1i+1} & \cdots & a_{1j} \\ a_{21} & \cdots & a_{2i-1} & 0 & a_{2i+2} & \cdots & a_{2j} \\ \vdots & & & \vdots & \vdots & & \\ a_{j1} & \cdots & a_{ji-1} & 1 & a_{ji+2} & \cdots & a_{jj} \end{bmatrix}}{\Delta_j} .$$

In the new basis the quadratic form is expressed as a sum of squares

$$z = \frac{1}{\Delta_1} \bar{x}_1^2 + \frac{\Delta_1}{\Delta_2} \bar{x}_2^2 + \dots + \frac{\Delta_{n-1}}{\Delta_n} \bar{x}_n^2 .$$

Note that a necessary and sufficient condition for z to be positive definite is

$$\Delta_i > 0 , \quad i = 1,2,\dots,n .$$

Sufficiency is clear, for if $\Delta_i > 0$ $(i = 1,2,\dots,n)$ then there is a basis in which the matrix of the quadratic form is diagonal with respect to this basis. Then

$$z = x^T \begin{bmatrix} \frac{1}{\Delta_1} & & & 0 \\ & \frac{\Delta_1}{\Delta_2} & & \\ & & \ddots & \\ 0 & & & \frac{\Delta_{n-1}}{\Delta_n} \end{bmatrix} x$$

$$= \overline{x}^T F^T \begin{bmatrix} \frac{1}{\Delta_1} & & & 0 \\ & \frac{\Delta_1}{\Delta_2} & & \\ & & \ddots & \\ 0 & & & \frac{\Delta_{n-1}}{\Delta_n} \end{bmatrix} F\overline{x}$$

$$= y^T \begin{bmatrix} \frac{1}{\Delta_1} & & & 0 \\ & \frac{\Delta_1}{\Delta_2} & & \\ & & \ddots & \\ 0 & & & \frac{\Delta_{n-1}}{\Delta_n} \end{bmatrix} y > 0 \quad \text{in any basis .}$$

Conversely, if z is positive definite then $\Delta_i > 0$ $(i = 1,2,...,n)$. Suppose that

$$\Delta_k = \begin{bmatrix} <e_1,Ae_1> & ... & <e_1,Ae_k> \\ \vdots & & \\ <e_k,Ae_1> & ... & <e_k,Ae_k> \end{bmatrix} = 0$$

for some k.

Then there are scalars $\mu_1,...,\mu_k$ not all zero such that

$$\mu_1 <e_1,Ae_i> + ... + \mu_k <e_k,Ae_i> = 0$$

for $i = 1,2,...,k$.

Thus

$$<\mu_1 e_1 + ... + \mu_k e_k, Ae_i> = 0 \quad \text{for} \quad i = 1,2,...,k,$$

and hence

$$<\mu_1 e_1 + ... + \mu_k e_k, A(\mu_1 e_1 + ... + \mu_k e_k)>$$

$$= (\mu_1 e_1 + ... + \mu_k e_k)^T A(\mu_1 e_1 + ... + \mu_k e_k) = 0 ,$$

where

$$\mu_1 e_1 + ... + \mu_k e_k \neq 0 ,$$

which contradicts the fact that $z = x^T Ax$ is positive definite. Therefore if $x^T Ax > 0$ then $\Delta_i \neq 0$ $(i = 1,...,n)$. But the fact that $\Delta_i \neq 0$ implies that the matrix of the quadratic form is diagonalizable, that is in some basis

$$z = \frac{1}{\Delta_1} \bar{x}_1^2 + \frac{\Delta_1}{\Delta_2} \bar{x}_2^2 + ... + \frac{\Delta_{n-1}}{\Delta_n} \bar{x}_n^2 .$$

Hence if the quadratic form is positive definite then all leading principal minors must be positive.

4. DIFFERENTIABILITY

Many of the techniques of calculus have as their foundation the idea of approximating a function by a linear function or by an *affine function.* A function

$$A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

is affine if there exists a linear function

$$L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

and a vector y_o in $\mathbb{R}^m$ such that

$$A(x) = L(x) + y_o$$

for every x in $\mathbb{R}^n$. Affine functions are the foundation of the differential calculus of vector functions. However, it is the linear part L of an affine function that is most useful, so we often speak of finding a *linear approximation.*

We shall analyze the possibility of approximating an arbitrary vector function near a point x_o of its domain by an affine function A. The general idea is the possibility of replacing near x_o what may be a very complicated function by an affine function that "best approximates" it. We begin by requiring that

$$f(x_o) = A(x_o) .$$

Since

$$A(x) = L(x) + y_o ,$$

where *L is linear,* we obtain $f(x_o) = Lx_o + y_o$, and so $y_o = -Lx_o + f(x_o)$.

Then the linearity of L shows that

$$A(x) = L(x - x_o) + f(x_o) .$$

This requirement is significant, but L could still be any linear function with the same domain and range as f. Some additional requirement is necessary. A natural condition, and the one we shall require, is that

$$f(x) - A(x)$$

approaches **0** faster than x approaches x_o. That is, we demand that

$$\lim_{x \to x_o} \frac{f(x) - f(x_o) - L(x - x_o)}{||x - x_o||} = 0 .$$

In addition, we want to guarantee that x can approach x_o from any direction; to do this we assume that x_o is an interior point of the domain f.

A function f : $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ will be called *differentiable* at x_o if [10]:

(i) x_o is an interior point of the domain of f.

(ii) there is an affine function that approximates f near x_o. That is, there exists a linear function L : $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ such that

$$\lim_{x \to x_o} \frac{f(x) - f(x_o) - L(x - x_o)}{||x - x_o||} = 0 .$$

The linear function L is called the *differential* of f at x_o. The function f is said to be *differentiable* if f is differentiable at every point of its domain.

In dimension 1, an affine function has the form ax + b. Hence, a real-valued function f(x) of a real variable x that is differentiable at x_o can be approximated near x_o by a function

$$A(x) = ax + b .$$

since $f(x_o) = A(x_o) = ax_o + b$, we obtain

$$A(x) = ax + b = a(x - x_o) + f(x_o) .$$

The linear part of A (denoted earlier by L) is in this case just multiplication by the real number a. The norm of a real; number is its absolute value, so condition (ii) of the definition of differentiability becomes

$$\lim_{x \to x_o} \frac{f(x) - f(x_o) - a(x - x_o)}{|x - x_o|} = 0$$

which is equivalent to

$$\lim_{x \to x_o} \frac{f(x) - f(x_o)}{x - x_o} = a \ .$$

The number a is commonly devoted by $f'(x_o)$ and is called the derivative of f at x_o. The affine function A is therefore given by

$$A(x) = f(x_o) + f'(x_o)\,(x - x_o)\ .$$

Its graph is the tangent line to the graph of f at x_o.

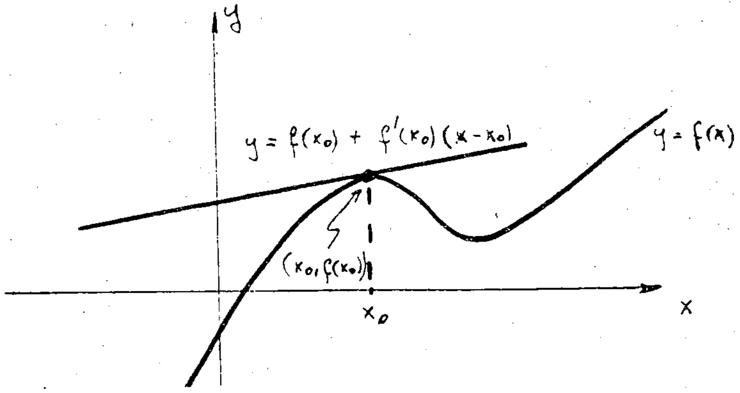


Fig. 4.1. Illustration of the notion of the derivative.

A linear function

$$L : \mathbb{R}^n \to \mathbb{R}^m$$

is representable by an m×n matrix. It follows that L is *uniquely* determined by f at each interior point of the domain of f. Thus we can speak of the differential of f at x_o, and of the function

$$A(x) = f(x_o) + L(x - x_o)$$

as the best affine approximation to f near x_o.

The Derivative Matrix

To find the matrix representation of the differential L of a function f from $\mathbb{R}^n$ to $\mathbb{R}^m$, we use the natural basis $\{e_1,...,e_m\}$ for the domain space $\mathbb{R}^n$. If x_o is an interior point of the domain of f, the vectors

$$x_j = x_o + t\, e_j \ , \ j = 1,...,n$$

are all in the domain of f for sufficiently small t. By condition (ii) of the definition of the differential, we have

$$\lim_{t\to 0} \frac{f(x_j) - f(x_o) - L(te_j)}{t} = 0$$

for $j = 1,...,n$. Since L is a linear function, this means that

$$\lim_{t\to 0} \frac{f(x_j) - f(x_o)}{t} = Le_j$$

for $j = 1,...,n$. But Le_j is the jth column of the matrix L. On the other hand, the vector x_j differs from x_o only in the jth coordinate, and in that coordinate the difference is just the number t. Therefore the left side of the last equation is precisely the partial derivative

$$(\partial f/\partial x_j)(x_o) .$$

Since vector limits are computed by taking the limit of each coordinate function, it follows immediately that if

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \quad \text{then} \quad \frac{\partial f}{\partial x_i}(x) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_i}(x) \\ \vdots \\ \dfrac{\partial f_m}{\partial x_i}(x) \end{bmatrix}.$$

Thus if the coordinate functions of f are $f_1,\ldots,f_m$, then

$$\frac{\partial f}{\partial x_j}(x_o) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_j}(x_o) \\ \vdots \\ \dfrac{\partial f_m}{\partial x_j}(x_o) \end{bmatrix},$$

and the entire matrix of L has the form

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(x_o) & \cdots & \dfrac{\partial f_1}{\partial x_n}(x_o) \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(x_o) & \cdots & \dfrac{\partial f_m}{\partial x_n}(x_o) \end{bmatrix}.$$

This matrix is called the *Jacobian matrix,* or *derivative matrix,* of f at x_o, and is denoted $f'(x_o)$; we sometimes simply refer to $f'(x_o)$ as the derivative of f at x_o. We can summarize what we have just proved as follows.

Theorem 4.1. If a function

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

is differentiable at x_o, then the differential of f at x_o is uniquely determined of f at x_o is uniquely determined and is represented by the derivative matrix $f'(x_o)$. The best affine approximation to f near x_o is then given by

$$A(x) = f(x_o) + f'(x_o)(x - x_o) .$$

The columns of the derivative matrix $f'(x_o)$ are vector partial derivatives. The vector partial

$$(\partial f/\partial x_j)(x_o)$$

is a tangent vector at $f(x_o)$ to the image curve of f obtained by varying only the jth coordinate variable x_j.

Gradient Vectors

If f is a differentiable real-valued function

$$f : R^n \longrightarrow R$$

then the function ∇f defined by

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1} (x) , \dots , \frac{\partial f}{\partial x_n} (x) \right]^T$$

is called the *gradient* of f. The gradient is evidently a function from $\mathbb{R}^n$ to $\mathbb{R}^n$, and it can be pictured as a *vector field,* that is, by drawing the arrow representing $\nabla f(x)$ so that its tail starts at f(x) instead of the origin. Physically, the direction and length of the arrow $\nabla f(x)$ can often be thought of as the direction and speed of a fluid flow at the point x to which the arrow is attached [10].

Chain Rule

Now we shall prove a chain rule for differentiating the composition $g(f(t))$, of a function $f: \mathbb{R} \rightarrow \mathbb{R}^n$ and a function $g: \mathbb{R}^n \rightarrow \mathbb{R}$.

Theorem 4.2. Let g be real-valued continuously differentiable on an open set D in $\mathbb{R}^n$ and let $f(t)$ be defined and differentiable for $a < t < b$, taking its values in D. Then the composite function $F(t) = g(f(t))$ is differentiable for $a < t < b$ and

$$F'(t) = \nabla^T g(f(t)) \cdot f'(t)$$

Proof [10]:

By definition

$$F'(t) = \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} =$$

$$= \lim_{h \to 0} \frac{g(f(t+h)) - g(f(t))}{h}$$

if the limit exists. Since f is differentiable, it is continuous. We now apply the mean-value theorem to g, getting

$$g(y) - g(x) = g'(x_o)(y - x) =$$

$$= \nabla^T g(x_o) \cdot (y - x)$$

where x_o is some point on the segment joining y and x. Letting $x = f(t)$ and $y = f(t+h)$, we have

$$\frac{F(t+h) - F(t)}{h} = \nabla^T g(x_o) \cdot \frac{f(t+h) - f(t)}{h} .$$

Thus

$$F'(t) = \lim_{h \to 0} \nabla^T g(x_o) \cdot \frac{f(t+h) - f(t)}{h} =$$

$$= \nabla^T g(f(t)) \cdot f'(t) .$$

Normal Vectors

The gradient is particularly useful in analyzing the level set of a real-valued function. Recall that a *level set* S of a function f is a set of points x satisfying

$$f(x) = k \quad \text{for some constant } k .$$

For $f : \mathbb{R}^2 \to \mathbb{R}$ we are usually interested in S when it is a curve, and for $f : \mathbb{R}^3 \to \mathbb{R}$, the sets S most often considered are surfaces.

To say that a point x_o is on the level S corresponding to level, k is to say that

$$f(x_o) = k .$$

Now suppose that there is a curve γ lying in S and parameterized by a continuously differentiable function

$$g : \mathbb{R} \to \mathbb{R}^n$$

Suppose also that $g(t_o) = x_o$ and

$$g'(t_o) = v \neq 0 ,$$

so that v is a tangent vector to γ at x_o

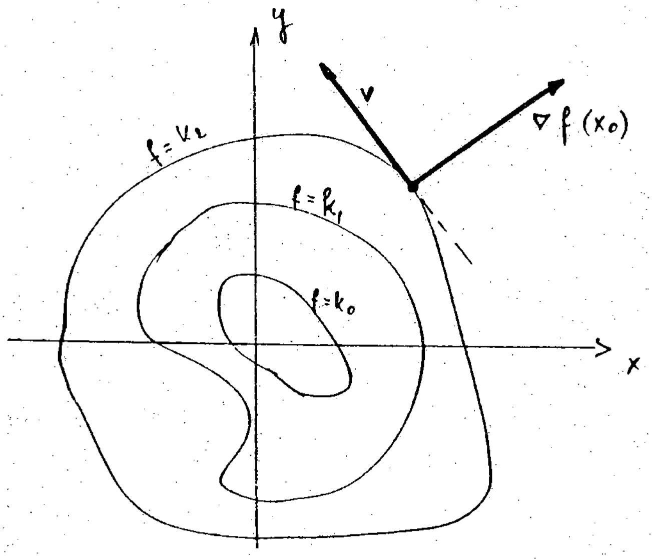


Fig. 4.2. Illustration of the level sets.

Applying the chain rule to the function

$$h(t) = f(g(t)) \quad \text{at } t_o$$

gives

$$h'(t_o) = \nabla^T f(g(t_o)) \cdot g'(t_o) =$$

$$= \nabla^T f(x_o) \cdot v$$

But since γ lies on S, we have

$$h(t) = f(g(t)) = k$$

that is, h is constant. Thus $h'(t_o) = 0$ and

$$\nabla^T f(x_o) \cdot v = 0$$

Hence we have proved, assuming f continuously differentiable, the following theorem.

Theorem 4.3. If $\nabla f(x_o)$ is not zero, then it is perpendicular to the tangent vector to an arbitrary smooth curve passing through x_o on the level set determined by $f(x) = k$. For this reason it is natural to say that $\nabla f(x_o)$ is perpendicular or normal to the level set S defined by $f(x) = k$ at x_o and to take as the tangent plane (or line) to S at x_o the set of all points x satisfying

$$\nabla^T f(x_o) \cdot (x - x_o) = 0 \quad \text{if}$$

$$\nabla f(x_o) \neq 0 \, .$$

We see that *the direction of maximum increase of a real-valued differentiable function at a point is perpendicular to the level set of the function through that point.* The reason is that $\nabla f(x_o)$ is the direction of maximum increase of f at x_o, and at the same time is perpendicular to the level set through x_o, determined by $f(x) = k$.

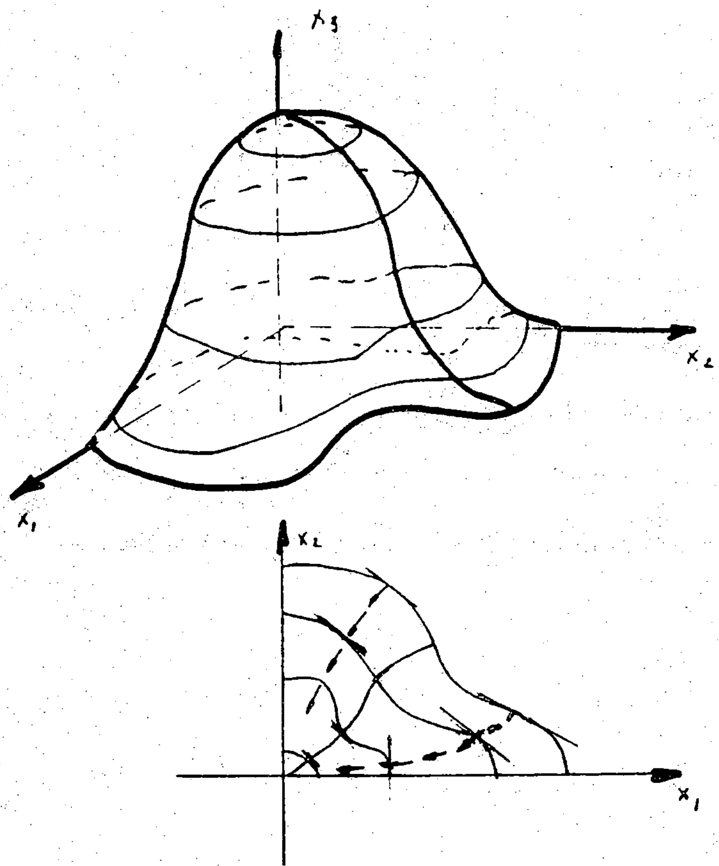


Fig. 4.3. Illustration of a path of steepest ascent.

The curve on the graph running from bottom to top has the property that its projection onto (x_1,x_2)-plane is always perpendicular to the level curves, and is called a *path of steepest ascent,* because it always heads in the direction of maximum increase for f.

REVIEW OF TAYLOR SERIES

The basis for many numerical methods and linear models for dynamic systems can be traced back to the Taylor formula. In what follows we discuss the Taylor series.

Let us assume that a function $f:\mathbb{R}^1 \to \mathbb{R}^1$ is n times differentiable on an interval $a \leq x \leq b$. Denote $h = b - a$. Then f(b) may be represented as follows

$$f(b) = f(a) + \frac{h}{1!} f'(a) + \frac{h^2}{2!} f''(a) + \ldots$$

$$+ \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n,$$

where

$$R_n = \frac{h^n(1-\Theta)^{n-1}}{(n-1)!} f^{(n)}(a + \Theta h) ,$$

and Θ is a suitably chosen number satisfying $0 < \Theta < 1$.

Proof [6]:

We have

$$R_n = f(b) - f(a) - \frac{h}{1!} f'(a) - \frac{h^2}{2!} f''(a) - \ldots - \frac{h^{n-1}}{(n-1)!} f^{n-1}(a) .$$

Denote by $g_n(x)$ an auxiliary function obtained from R_n by replacing a by x, hence

$$g_n(x) = f(b) - f(x) - \frac{b-x}{1!} f'(x)$$

$$- \frac{(b-x)^2}{2!} f''(x) - \ldots - \frac{(b-x)^{n-1}}{(n-1)!} f^{(n-1)}(x).$$

Differentiating $g_n(x)$ yields

$$g_n'(x) = -f'(x) + \left[f'(x) - \frac{b-x}{1!} f''(x)\right]$$

$$+ \left[2\frac{b-x}{2!}f''(x) - \frac{(b-x)^2}{2!} f'''(x)\right]$$

$$+ \ldots \left[(n-1)\frac{(b-x)^{n-2}}{(n-1)!} f^{(n-1)}(x)\right.$$

$$\left. - \frac{(b-x)^{n-1}}{(n-1)!} f^{(n)}(x)\right].$$

Note that

$$g_n'(x) = -\frac{(b-x)^{n-1}}{(n-1)!} f^{(n)}(x) .$$

Observe also that

$$g_n(b) = 0 \quad \text{and} \quad g_n(a) = R_n .$$

Applying the mean value theorem yields

$$\frac{g_n(b) - g_n(a)}{b-a} = g_n'(a + \Theta h) .$$

The above equation is equivalent to

$$-\frac{R_n}{h} = -\frac{(b - a - \Theta h)^{n-1}}{(n-1)!} f^{(n)}(a + \Theta h) .$$

$$= -\frac{h^{n-1}(1-\Theta)^{n-1}}{(n-1)!} f^{(n)} (a + \Theta h) .$$

Hence

$$R_n = \frac{h^n(1-\Theta)^{n-1}}{(n-1)!} f^{(n)}(a + \Theta h).$$

□

We say that a function is *analytic* at a point P if it can be expanded in a Taylor series about P.

A function is analytic in a region if it is analytic at every point of the region.

Note that the existence of all the derivatives of a function at a point does not necessarily imply that the function is analytic at this point.

Hence given an arbitrary analytic function f(x), the Taylor series for the function is

$$f(x) = f(x_0) + \sum_{k=1}^{\infty} \frac{(x - x_0)^k}{k!} f^{(k)}(x_0) .$$

We now turn to the Taylor series expansion about the point $x_0 \in \mathbb{R}^n$ of a real-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

Let x and x_0 be fixed vectors in $\mathbb{R}^n$ and let $z = x_0 + \alpha(x - x_0)$. Define $F : \mathbb{R}^1 \longrightarrow \mathbb{R}^1$ by ([1]):

$$F(\alpha) = f(z) = f(x_0 + \alpha(x - x_0)) .$$

Using the chain rule, we obtain

$$\frac{dF}{d\alpha} = \frac{df}{dz}\frac{dz}{dt} = \frac{df}{dz}(x - x_0)$$

$$= (x - x_0)^T \left(\frac{df}{dz}\right)^T$$

and

$$\frac{d^2F}{d\alpha^2} = \frac{d}{d\alpha}\left(\frac{dF}{d\alpha}\right) = (x - x_0)^T \frac{d}{d\alpha}\left(\frac{df}{dz}\right)^T$$

$$= (x - x_0)^T \frac{d}{dz}\left(\frac{df}{dz}\right)^T \frac{dz}{d\alpha}$$

$$= (x - x_0)^T \left(\frac{d^2 f}{dz^2} \right)^T (x - x_0)$$

$$= (x - x_0)^T \; \frac{d^2 f}{dz^2} \; (x - x_0) \; .$$

Observe that ([1])

$$f(x) = F(1) = F(0) + \frac{1}{1!} F'(0) + \frac{1}{2!} \; F''(0) + \ldots$$

Hence

$$f(x) = f(x_0) + \frac{1}{1!} \; f'(x_0)(x - x_0) + \frac{1}{2!} (x - x_0)^T f''(x_0)(x - x_0) + \ldots ,$$

where

$$f' = \frac{df}{dx} = \left[\frac{\partial f}{\partial x_1} \, , \, \frac{\partial f}{\partial x_2} \, , \, \ldots \, , \, \frac{\partial f}{\partial x_n} \right] ,$$

and

$$f'' = \left(\frac{d^2 f}{dx^2} \right)^T = \frac{d}{dx} \left(\frac{df}{dx} \right]^T$$

$$= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} .$$

REFERENCES

[1] K. G. Binmore, "*Calculus,*" Cambridge University Press, Cambridge, 1986.

[2] V. G. Boltyanskii, "*Mathematical Methods of Optimal Control,*" Holt, Rinehart and Winston, New York, 1971.

[3] V. F Dem'yanov and L. V. Vasilév, "*Nondifferentiable Optimization,*" Optimization Software, Inc., Publications Division, New York, 1985.

[4] I. M. Gel'fand, "*Lectures on Linear Algebra,*" Interscience Publishers, Inc., New York, 1961.

[5] V. N. Faddeeva, "*Computational Methods of Linear Algebra,*" Dover Publications, Inc., New York, 1959.

[6] K. Kuratowski, "*Introduction to Calculus,*" Second Edition, Pergamon Press, International Series of Monographs in Pure and Applied Mathematics, Vol. 17, Warszawa, 1969.

[7] D. G. Luenberger, "*Linear and Nonlinear Programming,*" Second Edition, Addison-Wesley, Reading, Mass., 1984.

[8] A. Mostowski, and M. Stark, "*Elements of Higher Algebra,*" PWN - Polish Scientific Publishers, Warszawa, 1958.

[9] S. L. Salas and E. Hille, "*Calculus: One and Several Variables,*" Fourth Edition, J. Wiley and Sons, New York, 1982.

[10] R. E. Williamson and H. F. Trotter, "*Multivariable Mathematics,*" Second Edition, Prentice-Hall, Englewood Cliffs, N.J., 1979.