Lecture1

Optimization

Introduction, Basic concepts

Vector Spaces and Matrices

We define a *column* n-vector to be an array of n numbers, denoted

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

The number a_i is called the ith component of the vector $\boldsymbol{a}$. Denote by $\mathbb{R}$ the set of real numbers, and by $\mathbb{R}^n$ the set of column n-vectors with real components. We call $\mathbb{R}^n$ an n-dimensional *real vector space*. We commonly denote elements of $\mathbb{R}^n$ by lower-case bold letters, e.g., $\boldsymbol{x}$. The components of $\boldsymbol{x} \in \mathbb{R}^n$ are denoted $x_1, \ldots, x_n$.

We define a *row* n-vector as

$$[a_1, a_2, \ldots, a_n].$$

The *transpose* of a given column vector $\boldsymbol{a}$ is a row vector with corresponding elements, denoted $\boldsymbol{a}^T$. Equivalently, we may write $\boldsymbol{a} = [a_1, a_2, \ldots, a_n]^T$.

In this course, the term vector (without the qualifier row or column) will be refer to “column vector”.

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

$$\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k = \boldsymbol{0}$$

implies that all coefficients α_i, $i = 1, \ldots, k$, are equal to zero. A set of the vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ is *linearly dependent* if it is not linearly independent.

- ❖ What about the set of composed of the single vector 0 ?

Note that the set composed of the single vector $\boldsymbol{0}$ is linearly dependent, for if $\alpha \neq 0$ then $\alpha \boldsymbol{0} = \boldsymbol{0}$. In fact, any set of vectors containing the vector $\boldsymbol{0}$ is linearly dependent.

- ❖ What about a set composed of the single nonzero vector?

A set composed of a single nonzero vector $\boldsymbol{a} \neq \boldsymbol{0}$ is linearly independent since $\alpha \boldsymbol{a} = \boldsymbol{0}$ implies $\alpha = 0$.

A vector $\boldsymbol{a}$ is said to be a *linear combination* of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ if there are scalars $\alpha_1, \ldots, \alpha_k$ such that

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

Proposition 2.1 *A set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is linearly dependent if and only if one of the vectors from the set is a linear combination of the remaining vectors.*

Proof. $\Rightarrow$: If $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is linearly dependent then

$$\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k = \boldsymbol{0},$$

where at least one of the scalars $\alpha_i \neq 0$, whence

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

$\Leftarrow$: Suppose

$$\boldsymbol{a}_1 = \alpha_2 \boldsymbol{a}_2 + \alpha_3 \boldsymbol{a}_3 + \cdots + \alpha_k \boldsymbol{a}_k,$$

then

$$(-1)\boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k = \boldsymbol{0}.$$

Because the first scalar is nonzero, the set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is linearly dependent. The same argument holds if $\boldsymbol{a}_i$, $i = 2, \ldots, k$, is a linear combination of the remaining vectors. ∎

A subset $\mathcal{V}$ of $\mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ if $\mathcal{V}$ is closed under the operations of vector addition and scalar multiplication. That is, if $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors in $\mathcal{V}$, then the vectors $\boldsymbol{a} + \boldsymbol{b}$ and $\alpha\boldsymbol{a}$ are also in $\mathcal{V}$ for every scalar α.

Every subspace contains the zero vector $\mathbf{0}$, for if $\boldsymbol{a}$ is an element of the subspace, so is $(-1)\boldsymbol{a} = -\boldsymbol{a}$. Hence, $\boldsymbol{a} - \boldsymbol{a} = \mathbf{0}$ also belongs to the subspace.

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ be arbitrary vectors in $\mathbb{R}^n$. The set of all their linear combinations is called the *span* of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ and is denoted

$$\operatorname{span}[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k] = \left\{ \sum_{i=1}^{k} \alpha_i \boldsymbol{a}_i : \alpha_1, \ldots, \alpha_k \in \mathbb{R} \right\} .$$

Given a vector $\boldsymbol{a}$, the subspace $\operatorname{span}[\boldsymbol{a}]$ is composed of the vectors $\alpha\boldsymbol{a}$, where α is an arbitrary real number ($\alpha \in \mathbb{R}$). Also observe that if $\boldsymbol{a}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ then

$$\operatorname{span}[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k, \boldsymbol{a}] = \operatorname{span}[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k].$$

The span of any set of vectors is a subspace.

Given a subspace $\mathcal{V}$, any set of linearly independent vectors $\{a_1, a_2, \ldots, a_k\} \subset \mathcal{V}$ such that $\mathcal{V} = \text{span}[a_1, a_2, \ldots, a_k]$ is referred to as a *basis* of the subspace $\mathcal{V}$. All bases of a subspace $\mathcal{V}$ contain the same number of vectors. This number is called the *dimension* of $\mathcal{V}$, denoted $\dim \mathcal{V}$.

Proposition 2.2 *If $\{a_1, a_2, \ldots, a_k\}$ is a basis of $\mathcal{V}$, then any vector a of $\mathcal{V}$ can be represented uniquely as*

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_k a_k,$$

where $\alpha_i \in \mathbb{R}$, $i = 1, 2, \ldots, k$.

Proof. To prove the uniqueness of the representation of $\boldsymbol{a}$ in terms of the basis vectors, assume that

$$\boldsymbol{a} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k$$

and

$$\boldsymbol{a} = \beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2 + \cdots + \beta_k \boldsymbol{a}_k.$$

We now show that $\alpha_i = \beta_i$, $i = 1, \ldots, k$. We have

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

or

$$(\alpha_1 - \beta_1)\boldsymbol{a}_1 + (\alpha_2 - \beta_2)\boldsymbol{a}_2 + \cdots + (\alpha_k - \beta_k)\boldsymbol{a}_k = \boldsymbol{0}.$$

Because the set $\{\boldsymbol{a}_i : i = 1, 2, \ldots, k\}$ is linearly independent, $\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \cdots = \alpha_k - \beta_k = 0$, that is, $\alpha_i = \beta_i$, $i = 1, \ldots, k$. ∎

Suppose we are given a basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ of $\mathcal{V}$ and a vector $\boldsymbol{a} \in \mathcal{V}$ such that

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

The coefficients α_i, $i = 1, \ldots, k$, are called the *coordinates* of $\boldsymbol{a}$ with respect to the basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$.

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

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \ldots, \quad \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The reason for calling these vectors the natural basis is that

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \boldsymbol{e}_1 + x_2 \boldsymbol{e}_2 + \cdots + x_n \boldsymbol{e}_n.$$

Rank of a matrix

A *matrix* is a rectangular array of numbers, commonly denoted by upper-case bold letters, e.g., $\boldsymbol{A}$. A matrix with m rows and n columns is called an $m \times n$ matrix, and we write

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

Let us denote the kth column of $\boldsymbol{A}$ by $\boldsymbol{a}_k$, that is,

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

The maximal number of linearly independent columns of $\boldsymbol{A}$ is called the *rank* of the matrix $\boldsymbol{A}$, denoted rank $\boldsymbol{A}$. Note that rank $\boldsymbol{A}$ is the dimension of span$[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$.

A pth-order *minor* of an $m \times n$ matrix $\boldsymbol{A}$, with $p \leq \min(m, n)$, is the determinant of a $p \times p$ matrix obtained from $\boldsymbol{A}$ by deleting $m - p$ rows and $n - p$ columns.

If a matrix $\boldsymbol{A}$ has an rth-order minor $|\boldsymbol{M}|$ with the properties (i) $|\boldsymbol{M}| \neq 0$ and (ii) any minor of $\boldsymbol{A}$ that is formed by adding a row and a column of $\boldsymbol{A}$ to $\boldsymbol{M}$ is zero, then

$$\operatorname{rank} \boldsymbol{A} = r.$$

Thus, the rank of a matrix is equal to the highest order of its nonzero minor(s).

A *nonsingular* (or *invertible*) matrix is a square matrix whose determinant is nonzero.

Suppose that $\boldsymbol{A}$ is an $n \times n$ square matrix. Then, $\boldsymbol{A}$ is nonsingular if and only if there is another $n \times n$ matrix $\boldsymbol{B}$ such that

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A} = \boldsymbol{I}_n,$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ *identity matrix*:

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

We call the above matrix $\boldsymbol{B}$ the *inverse matrix* of $\boldsymbol{A}$, and write $\boldsymbol{B} = \boldsymbol{A}^{-1}$.

Consider the $m \times n$ matrix

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

The *transpose* of $\boldsymbol{A}$, denoted $\boldsymbol{A}^T$, is the $n \times m$ matrix

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix},$$

that is, the columns of $\boldsymbol{A}$ are the rows of $\boldsymbol{A}^T$, and vice versa. A matrix $\boldsymbol{A}$ is *symmetric* if $\boldsymbol{A} = \boldsymbol{A}^T$.

Inner Products and Norms

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

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

The following formulas hold:

1. $|a| = |-a|$;
2. $-|a| \leq a \leq |a|$;
3. $|a+b| \leq |a| + |b|$;
4. $||a| - |b|| \leq |a-b| \leq |a| + |b|$;
5. $|ab| = |a||b|$;
6. $|a| \leq c$ and $|b| \leq d$ imply $|a+b| \leq c+d$;
7. The inequality $|a| < b$ is equivalent to $-b < a < b$ (i.e., $a < b$ and $-a < b$). The same holds if we replace every occurrence of "$<$" by "$\leq$."
8. The inequality $|a| > b$ is equivalent to $a > b$ or $-a > b$. The same holds if we replace every occurrence of "$>$" by "$\geq$."

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we define the *Euclidean inner product* by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^T \boldsymbol{y}.$$

The inner product is a real-valued function $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ having the following properties:

1. Positivity: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$;
2. Symmetry: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$;
3. Additivity: $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$;
4. Homogeneity: $\langle r\boldsymbol{x}, \boldsymbol{y} \rangle = r\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for every $r \in \mathbb{R}$.

The properties of additivity and homogeneity in the second vector also hold, that is,

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle, \\ \langle \boldsymbol{x}, r\boldsymbol{y} \rangle &= r\langle \boldsymbol{x}, \boldsymbol{y} \rangle \quad \text{for every } r \in \mathbb{R}. \end{aligned}$$

The vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be *orthogonal* if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

The *Euclidean norm* of a vector $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\boldsymbol{x}^T \boldsymbol{x}}.$$

The Euclidean norm of a vector $\|\boldsymbol{x}\|$ has the following properties:

1. Positivity: $\|\boldsymbol{x}\| \geq 0$, $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$;
2. Homogeneity: $\|r\boldsymbol{x}\| = |r| \|\boldsymbol{x}\|$, $r \in \mathbb{R}$;
3. Triangle Inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

The Euclidean norm is an example of a general *vector norm*, which is any function satisfying the above three properties of positivity, homogeneity, and triangle inequality. Other examples of vector norms on $\mathbb{R}^n$ include the 1-norm, defined by $\|\boldsymbol{x}\|_1 = |x_1| + \cdots + |x_n|$, and the ∞-norm, defined by $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$. The Euclidean norm is often referred to as the 2-norm, and denoted $\|\boldsymbol{x}\|_2$. The above norms are special cases of the p-norm, given by

$$\|\boldsymbol{x}\|_p = \begin{cases} (|x_1|^p + \cdots + |x_n|^p)^{1/p} & \text{if } 1 \leq p < \infty \\ \max(|x_1|, \ldots, |x_n|) & \text{if } p = \infty \end{cases}.$$

Eigenvalues and Eigenvectors

Let $\boldsymbol{A}$ be an $n \times n$ square matrix. A scalar λ (possibly complex) and a nonzero vector $\boldsymbol{v}$ satisfying the equation $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ are said to be, respectively, an *eigenvalue* and *eigenvector* of $\boldsymbol{A}$. For λ to be an eigenvalue it is necessary and sufficient for the matrix $\lambda\boldsymbol{I} - \boldsymbol{A}$ to be singular, that is, $\det[\lambda\boldsymbol{I} - \boldsymbol{A}] = 0$, where $\boldsymbol{I}$ is the $n \times n$ identity matrix. This leads to an nth-order polynomial equation

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

We call the polynomial $\det[\lambda\boldsymbol{I} - \boldsymbol{A}]$ the *characteristic polynomial* of the matrix $\boldsymbol{A}$, and the above equation the *characteristic equation*. According to the fundamental theorem of algebra, the characteristic equation must have n (possibly nondistinct) roots that are the eigenvalues of $\boldsymbol{A}$.

Concepts from Geometry

Line segments

The *line segment* between two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ is the set of points on the straight line joining points $\boldsymbol{x}$ and $\boldsymbol{y}$ (see Figure 4.1). Note that if $\boldsymbol{z}$ lies on the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$, then

$$\boldsymbol{z} - \boldsymbol{y} = \alpha(\boldsymbol{x} - \boldsymbol{y}),$$

where α is a real number from the interval $[0, 1]$. The above equation can be rewritten as $\boldsymbol{z} = \alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}$. Hence, the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be represented as

$$\{\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y} : \alpha \in [0, 1]\}.$$

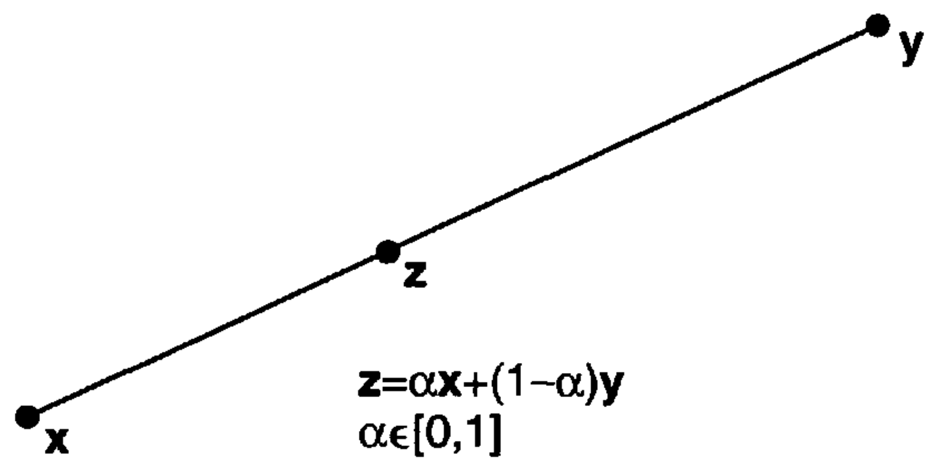

Figure 4.1 A line segment

Hyperplanes

Let $u_1, u_2, \ldots, u_n, v \in \mathbb{R}$, where at least one of the u_i is nonzero. The set of all points $\boldsymbol{x} = [x_1,\ x_2, \ldots, x_n]^T$ that satisfy the linear equation

$$u_1 x_1 + u_2 x_2 + \cdots + u_n x_n = v$$

is called a *hyperplane* of the space $\mathbb{R}^n$. We may describe the hyperplane by

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T \boldsymbol{x} = v\},$$

where

$$\boldsymbol{u} = [u_1, u_2, \ldots, u_n]^T.$$

A hyperplane is not necessarily a subspace of $\mathbb{R}^n$ since, in general, it does not contain the origin. For $n = 2$, the equation of the hyperplane has the form $u_1 x_1 + u_2 x_2 = v$, which is the equation of a straight line. Thus, straight lines are hyperplanes in $\mathbb{R}^2$. In $\mathbb{R}^3$ (three-dimensional space), hyperplanes are ordinary planes. By translating a hyperplane so that it contains the origin of $\mathbb{R}^n$, it becomes a subspace of $\mathbb{R}^n$ (see Figure 4.2). Because the dimension of this subspace is $n - 1$, we say that the hyperplane has dimension $n - 1$.

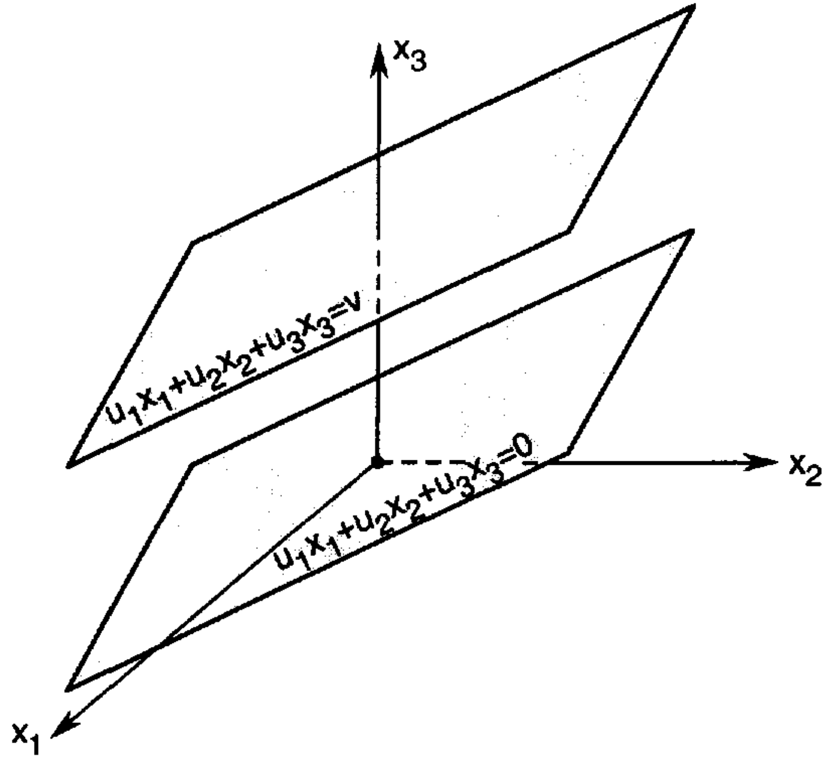

Figure 4.2 Translation of a hyperplane

The hyperplane $H = \{\boldsymbol{x} : u_1x_1 + \cdots + u_nx_n = v\}$ divides $\mathbb{R}^n$ into two *half-spaces*. One of these half-spaces consists of the points satisfying the inequality $u_1x_1 + u_2x_2 + \cdots + u_nx_n \geq v$, denoted

$$H_+ = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T\boldsymbol{x} \geq v\},$$

where, as before,

$$\boldsymbol{u} = [u_1, u_2, \ldots, u_n]^T.$$

The other half-space consists of the points satisfying the inequality $u_1x_1 + u_2x_2 + \cdots + u_nx_n \leq v$, denoted

$$H_- = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T\boldsymbol{x} \leq v\}.$$

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

Let $\boldsymbol{a} = [a_1, a_2, \ldots, a_n]^T$ be an arbitrary point of the hyperplane H. Thus, $\boldsymbol{u}^T\boldsymbol{a} - v = 0$. We can write

$$\begin{aligned} \boldsymbol{u}^T\boldsymbol{x} - v &= \boldsymbol{u}^T\boldsymbol{x} - v - (\boldsymbol{u}^T\boldsymbol{a} - v) \\ &= \boldsymbol{u}^T(\boldsymbol{x} - \boldsymbol{a}) \\ &= u_1(x_1 - a_1) + u_2(x_2 - a_2) + \cdots + u_n(x_n - a_n) = 0. \end{aligned}$$

The numbers $(x_i - a_i)$, $i = 1, \ldots, n$, are the components of the vector $\boldsymbol{x} - \boldsymbol{a}$. Therefore, the hyperplane H consists of the points $\boldsymbol{x}$ for which $\langle \boldsymbol{u}, \boldsymbol{x} - \boldsymbol{a} \rangle = 0$. In other words, the hyperplane H consists of the points $\boldsymbol{x}$ for which the vectors $\boldsymbol{u}$ and $\boldsymbol{x} - \boldsymbol{a}$ are orthogonal (see Figure 4.3). We call the vector $\boldsymbol{u}$ the *normal* to the hyperplane H. The set H_+ consists of those points $\boldsymbol{x}$ for which $\langle \boldsymbol{u}, \boldsymbol{x} - \boldsymbol{a} \rangle \geq 0$, and H_- consists of those points $\boldsymbol{x}$ for which $\langle \boldsymbol{u}, \boldsymbol{x} - \boldsymbol{a} \rangle \leq 0$.

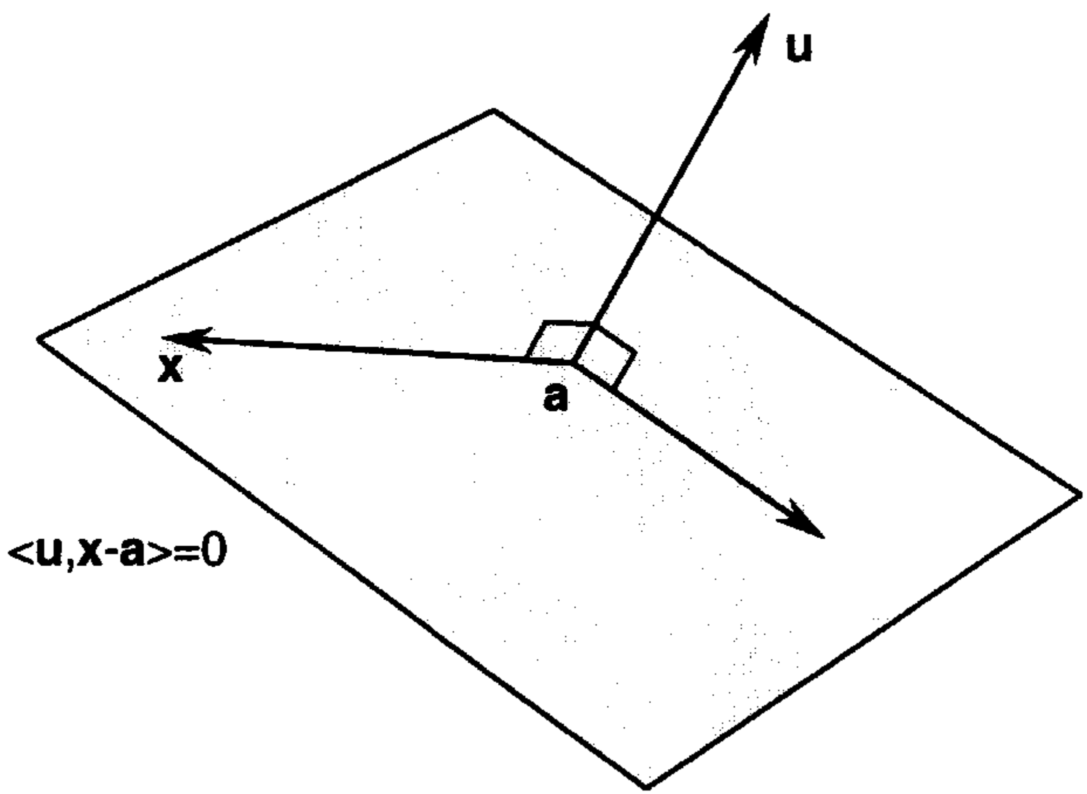

Figure 4.3 The hyperplane $H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T(\boldsymbol{x} - \boldsymbol{a}) = 0\}$

Convex Sets

Recall that the line segment between two points $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ is the set $\{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}, \alpha \in [0,1]\}$. A point $\boldsymbol{w} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$ (where $\alpha \in [0,1]$) is called a *convex combination* of the points $\boldsymbol{u}$ and $\boldsymbol{v}$.

A set $\Theta \subset \mathbb{R}^n$ is *convex* if for all $\boldsymbol{u}, \boldsymbol{v} \in \Theta$, the line segment between $\boldsymbol{u}$ and $\boldsymbol{v}$ is in Θ. Figure 4.4 gives examples of convex sets, whereas Figure 4.5 gives examples of sets that are not convex. Note that Θ is convex if and only if $\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v} \in \Theta$ for all $\boldsymbol{u}, \boldsymbol{v} \in \Theta$ and $\alpha \in (0,1)$.

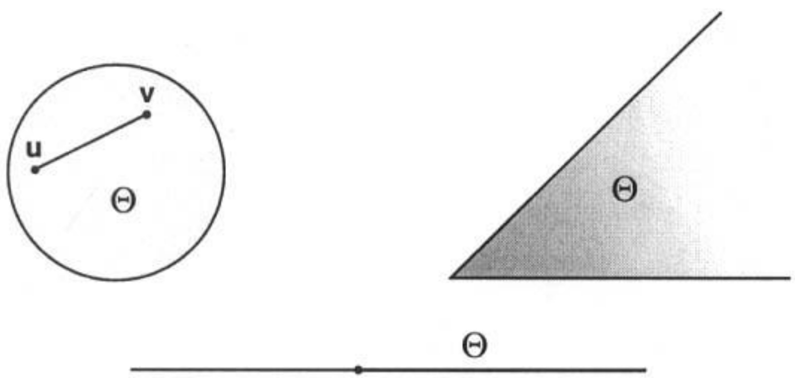

Figure 4.4 Convex sets

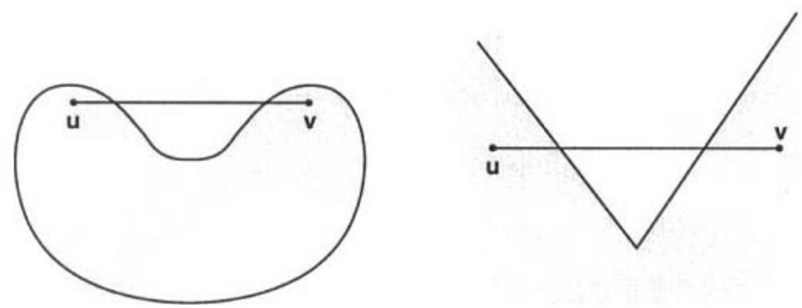

Figure 4.5 Sets that are not convex

Examples of convex sets include:

- the empty set
- a set consisting of a single point
- a line or a line segment
- a subspace
- a hyperplane
- a half-space
- $\mathbb{R}^n$.

A point $\boldsymbol{x}$ in a convex set Θ is said to be an *extreme point* of Θ if there are no two distinct points $\boldsymbol{u}$ and $\boldsymbol{v}$ in Θ such that $\boldsymbol{x} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$ for some $\alpha \in (0, 1)$. For example, in Figure 4.4, any point on the boundary of the disk is an extreme point, the vertex (corner) of the set on the right is an extreme point, and the endpoint of the half-line is also an extreme point.

Neighborhoods

A *neighborhood* of a point $\boldsymbol{x} \in \mathbb{R}^n$ is the set

$$\{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\},$$

where ε is some positive number. The neighborhood is also called the *ball* with radius ε and center $\boldsymbol{x}$.

In the plane $\mathbb{R}^2$, a neighborhood of $\boldsymbol{x} = [x_1, x_2]^T$ consists of all the points inside of a disc centered at $\boldsymbol{x}$. In $\mathbb{R}^3$, a neighborhood of $\boldsymbol{x} = [x_1, x_2, x_3]^T$ consists of all the points inside of a sphere centered at $\boldsymbol{x}$ (see Figure 4.7).

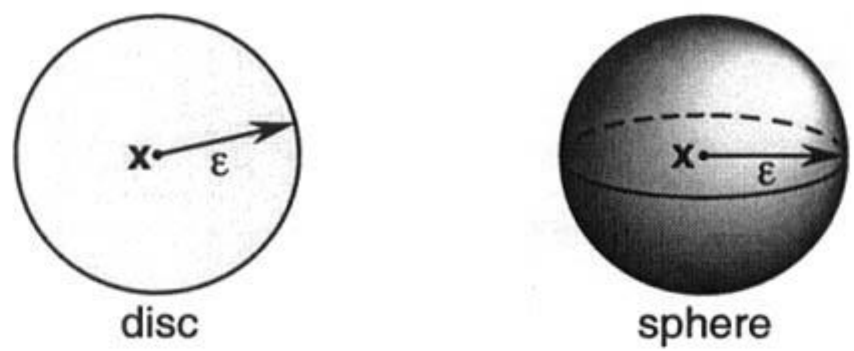

Figure 4.7 Examples of neighborhoods of a point in $\mathbb{R}^2$ and $\mathbb{R}^3$

A point $x \in S$ is said to be an *interior point* of the set S if the set S contains some neighborhood of x, that is, if all points within some neighborhood of x are also in S (see Figure 4.8). The set of all the interior points of S is called the *interior* of S.

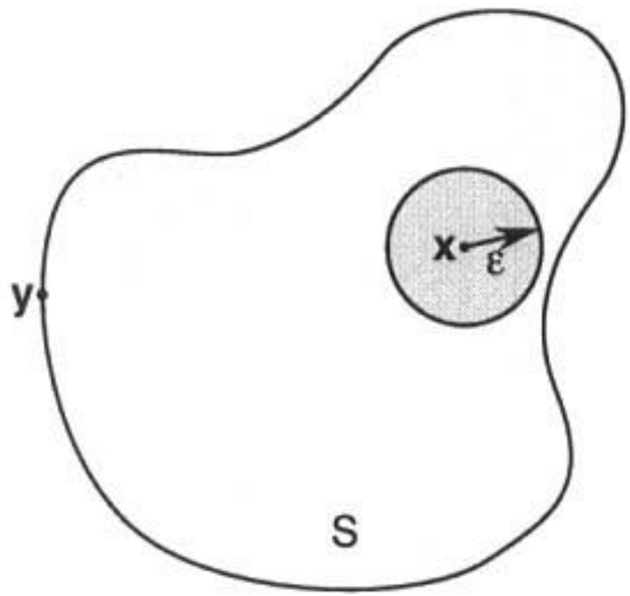

Figure 4.8 x is an interior point, while y is a boundary point

A point x is said to be a *boundary point* of the set S if every neighborhood of x contains a point in S and a point not in S (see Figure 4.8). Note that a boundary point of S may or may not be an element of 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, that is, if each of its points is an interior point, or equivalently, if S contains no boundary points.

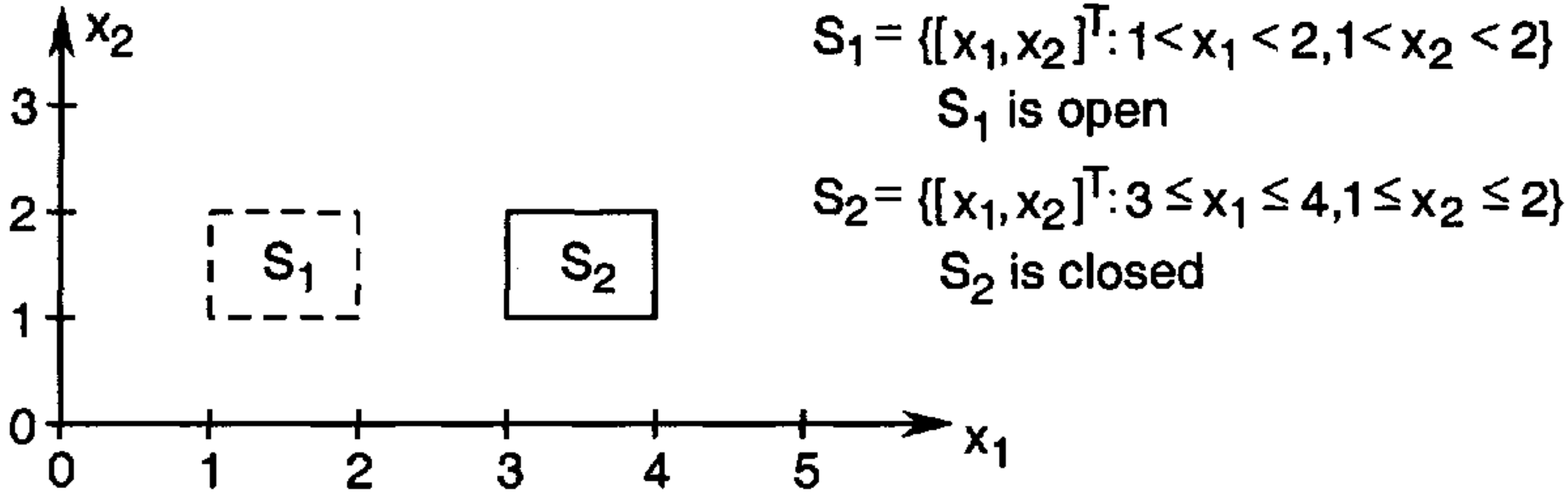

Figure 4.9 Open and closed sets

A set S is said to be *closed* if it contains its boundary (see Figure 4.9). We can show that a set is closed if and only if its complement is open.

A set that is contained in a ball of finite radius is said to be bounded.

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

Reference: Chong, E. K., & Zak, S. H. (2013). *An introduction to optimization* (Vol. 76). John Wiley & Sons, Chapter 1, Chapter 2, Chapter 4.